Fall 9-24-2020

Estimating School-Level Achievement in Belize

Betty Jean Usher-Tate
University of Nebraska-Lincoln, ushertate@huskers.unl.edu

ESTIMATING SCHOOL-LEVEL ACHIEVEMENT IN BELIZE

by

Betty Jean L. Usher-Tate

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Educational Psychology

(Quantitative, Qualitative and Psychometric Methods)

Under the Supervision of Professor Kurt F. Geisinger

Lincoln, Nebraska

September, 2020

ESTIMATING SCHOOL-LEVEL ACHIEVEMENT IN BELIZE

Betty Jean Usher-Tate, Ph.D.

University of Nebraska, 2020

Advisor: Kurt F. Geisinger

This dissertation consists of five chapters: introduction, literature review, methodology, results, and discussion with final thoughts. The research design of this dissertation study attended to structures, cultures, and characteristics associated with, and specific to, the Belize education system. The processes for data collection and types of analyses were appropriate, yielded meaningful results, and served as a segue for national application. The *Belize Educator Survey* was developed to capture the educators' voices and illuminate their relationship to educational achievement in Belize. The *Belize Educator Survey* was piloted and revised with direct input from educators and experts who work in the Belize education system. In this dissertation study, the *Belize Educator Survey* was the primary data collection instrument for a district-wide, full-coverage survey approach in one of the six districts in Belize. All primary school educators in the district were invited to participate (N = 524), and the response rate was 60.11% (n = 315).

One of the end goals of this study was to create a pragmatic way of estimating School-Level Achievement that incorporated data about all grade levels and include the voice of all educators associated with the school. Two methods, the Weighted-Indicator Estimation Protocol (WISP) and a Multilevel Achievement Estimation Protocol (MAEP), were compared. The Weighted-Indicator Scores Protocol estimation uses a combination of classic statistical analyses, while the Multilevel Achievement Estimation Protocol

version relies on a multilevel approach that estimates the within- and between-school statistics simultaneously. The 2018 Primary School Examinations school-level performance served as the primary criterion-referenced variable. Analyses of variances, comparisons of standard errors of the mean, and rank-order matchings show that the school-level estimates derived from the Weighted-Indicator Scores Protocol and Multilevel Achievement Estimation Protocol methods are statistically different.

Possible extensions of this study could identify or develop additional measures of academic performance that align to grade-level expectations in the Belize education system. The statistically significant negative associations between educators' perceptions, School Size and Location are also worthy of investigation for strategic planning purposes.

Acknowledgement

Dr. Geisinger and my home away from home family at the Buros Center for Testing have played an immeasurable role in this achievement. My committee members have also stood by me on this long journey: Dr. Bruning, Dr. DeAyala, Dr. Geisinger, Dr. Jones, and Dr. Lemanowski. A world of gratitude to Anthony, Avian, Bermuda, Devin, and my parents who rose to many challenges to carry me through this decade. To everyone who has touched my life positively in this academic sojourn; thank you.

It takes a whole village to raise a person, but one idiot to destroy him or her, or the village for that matter. I am sincerely thankful for all my professors, siblings, family, friends, and doctors who kept the idiots at bay – real and imagined. Amber, Blanche, Carolyn, Harry, Judith, Kristin, Lori, and Vie, I am uniquely indebted to each of you. There are so many people I would like to individually acknowledge, but that list might be a book in its own right; you are my collective village.

Dedicated to the cherished memories of Granny Winnie, Grandpa Nolan, Johnny, Aunt Audrey, Na, and Lana who I lost on this journey.

Table of Contents

Chapter 1

This chapter provides background information in addition to the purpose, limitations, and significance of the present study. This study addressed an area of interest in the Belizean education system, which is to assess School-Level Achievement (C. Babb, personal communication, June 20, 2018). To estimate School-Level Achievement in Belize, this research considered the selected variables associated with the education process, the educators, the schools, and the sources of student assessment data.

The Belize constitution defines the protection of inalienable rights and equality such that "equal protection should be given to children regardless of their social status, and that a just system should be ensured to provide for education and health on the basis of equality" (Government of Belize, 2011, p. 7). On November 20, 1989, the United Nations adopted the international treaty on the Convention on the Rights of the Child (CRC), which outlined global standards that ensure the protection, survival, and development of all children without discrimination (Human Rights Watch, 2014). On the 25th anniversary of its adoption, 194 countries, including Belize, were signatories to CRC. The three United Nations member countries that had not completed the ratification at that time were Somalia, South Sudan, and the United States of America (signed but not ratified). The CRC bore relevance to the present study because the treaty postulates education as a "right," not as a "privilege," therefore, Belize, as a signatory country, is accountable for the commitments made for education.

A seemingly universal goal in education is to prepare students to function in and contribute to the communities in which they live. However, communities are dynamic, so education systems occasionally need to be revised. One such example is the Partnership for 21st Century Learning (P21®) that embodied systemic changes to account for the advancement, accessibility, and versatility of technology. P21® embraces the technological advances and social mobility of this century. P21® emerged as a collaborative effort that drew from both business and education fields to develop a framework that addresses education’s role in globalization (P21®, 2018).

One of the resounding principles of the P21® Framework is to prepare students to function and innovate as global citizens in knowledge-based societies. Considerations for student preparation are geared toward life, career, use of technology, collaboration, and the pursuit of higher education (Ananiadou & Claro, 2009; Binkley et al., 2012; Jerald, 2009; Jhurree, 2005; Trilling & Fadel, 2009; Vockley, 2007). P21® as a framework may not be applicable in its entirety, but in principle, it is a useful model for systematic curricular efforts to meet the demands of changing needs. Belize is a developing country, but it is not devoid of technological advances available in developed countries. Belize shares the common economic reality of developing countries: few or no computers in primary schools, few infrastructural and technical resources, and limited funding (Jhurree, 2005). This economic reality is even more pronounced in rural schools. The economic reality of developing countries presents a plethora of disadvantages or opportunities to innovate, depending on one’s perspective as an educator, economist, or researcher.

International studies often identify and develop indicators for educational achievement that enable comparisons between and within participating countries (NCES, 2017; Stephens et al., 2015). Numerous published comparative studies collect education-related data of such magnitude. Examples include the International Association for the Evaluation of Educational Achievement's (IEA) Trends in International Mathematics and Science Study (TIMSS), the Progress in International Reading and Literacy Study (PIRLS), the Organization for Economic Co-operation and Development's (OECD) Program for International Student Assessment (PISA), and the Teaching and Learning International Survey (TALIS). These large-scale studies produced data that are used to identify educational indicators that support the comparison of education systems. Furthermore, the recognition of strong correlations between academic achievement and socioeconomic development continues to inspire international investigations of issues and indicators related to achievement (OECD, 2011). Some of these indicators include population demographics, enrollment in formal education, academic performance in specific content or subject areas, financial investment in education, and graduation rates. However, the integrity of the indicators, as mentioned earlier, was grounded in data collected in 20 industrialized nations referred to as the G-20 countries (NCES, 2017; Stephens et al., 2015). Notably, Belize does not share membership or the profile characteristics for population and development status of the countries in the G-20 study.

The most likely reasons why developing nations do not typically participate in international studies can be summed up as a combination of challenges of financial constraints, population size, insufficient resources, membership status, and politics. However, policymaking and health sciences studies indicate that developing countries can still benefit from information unveiled in studies conducted in other countries (Dolowitz & David, 1996; Lavis et al., 2004). Similarly, nonparticipating countries can benefit from international studies about education. There is value in teaching and learning from history.

Countries, like students, can be inspired by or better understand a phenomenon based on the experiences of others without having to “reinvent the wheel.” In this instance, the task at hand included identifying variables to use to estimate School-Level Achievement. One issue to consider when using past studies is the relevance to the current context. Over time and in response to research, the concept of educational achievement has changed. As noted, conceptualizations of educational achievement or “how we view and characterize achievement—affect what teachers teach and how they teach it, as well as what educational researchers study, and how they study it” (Cole, 1990, p. 2).

Meaningful estimates of school-level educational achievement are those that take into account systemic and socio-educational factors and can be used to stimulate positive reform when combined with appropriate interventions (Masinoa & Niño-Zarazúa, 2016). Research publications about educational achievement in Belize are limited. Therefore, findings from previous studies conducted in other parts of the world were necessary to inform design and gauge applicability for Belize within the present study. Within the context of this study, School-Level Achievement was conceptualized as a reference to the school’s overall performance, and it was not limited to a performance on a standardized test. This concept of School-Level Achievement does not ignore the expectations for satisfactory outcomes (academic and nonacademic) or the relationships directly associated with the school. This holistic concept of School-Level Achievement was construed as a multifaceted status, a summative reflection of the whole school as illustrated in Figure 1.

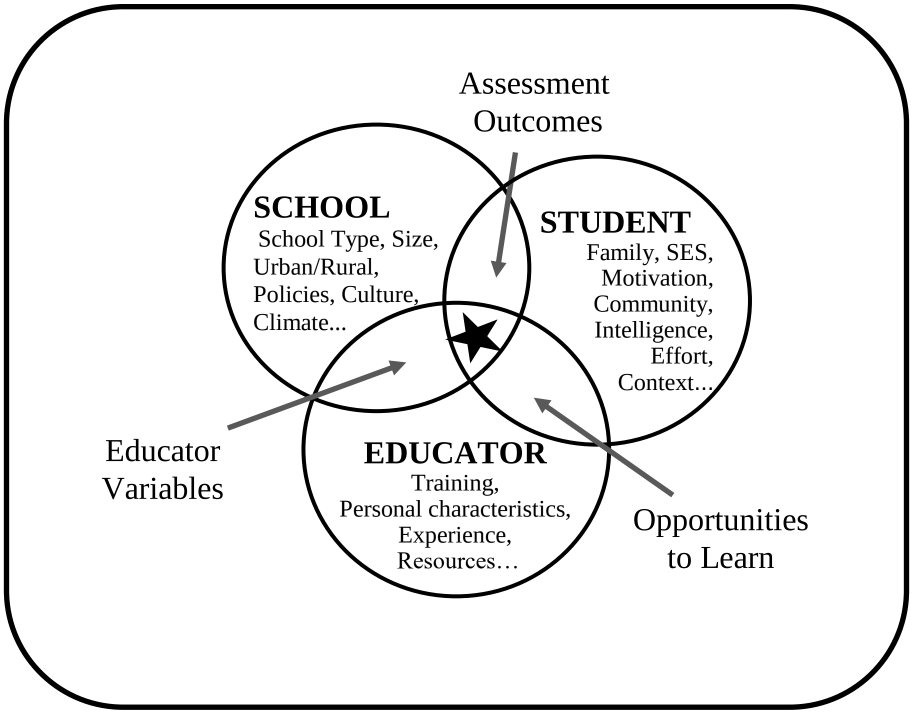


Figure 1. The conceptual model for educational achievement has overlapping circles that illustrate relationships between the school, the educator, and the student. The point where the roles, contributions, and influences converge is conceptualized as the core of educational achievement. The white space outside of the circles is the infinite number and combinations of unspecified possibilities or unknown factors that may or may not influence educational achievement in a meaningful way.

There was also a gap in the research literature about educational achievement in countries with relatively small populations, in developing countries, or in countries in the Caribbean. The education reports available for Belize are generally limited to descriptive statistics (sums, means, and percentages) that were published by educational entities in Belize (Belize Ministry of Education, 2007; Belize Ministry of Education, 2012c; Statistical Institute of Belize, 2013). Subsequently, the body of statistics reported in international statistics for Belize is typically descriptive and does not address complex issues relevant to education in Belize. The present research study was conducted with the intent to fill a part of this gap in educational research for Belize.

Statement of the Problem

In Belize, there were no preexisting models for estimating educational achievement at the school level. There was a need for a methodology protocol to estimate School-Level Achievement that was cost-effective and appropriate for small samples or small populations. Most importantly, the product needed to make sense to educators and be pragmatic (holistic, relevant, and replicable) for the type of education system in Belize.

The two drawbacks for addressing this problem are instrumentation and methodology that are appropriate for relatively small sample sizes. In Belize, there are no readily available or preexisting data banks for estimating educational achievement. There is also a need for comprehensive instruments to collect relevant data pertaining to selected variables that affect the education process and educational achievement. A primary research component of the present study pertains to putting together protocols and methods for estimating a multifaceted concept of educational achievement. The unique context of the education system in Belize was not found or described in other studies. Consequently, an instrument that is replicable within the Belizean framework was developed and pretested to collect data for the present study. In addition to determining what factors and variables are most relevant, the present study must also consider the methodological implications and make provisions for analyzing data derived from a relatively small and diverse population of educators and students.

English is the official language of Belize and the language of instruction and assessment in schools. However, English is not the first or home language of the majority of Belize's population (Statistical Institute of Belize, 2013). Home languages in Belize are generally associated with one's home culture, or the lingua franca (common language used across ethnic groups in a region) of the immediate community. There are instances of families who, for various reasons, do not pass down the language of the culture with which they identify. Nevertheless, the nomenclature (naming conventions) used in Belize to identify home languages is grouped as presented in the following list. References are to the people, culture, or ethnicity, and the corresponding languages are in parentheses.

- Creole (*Kriol*)
- Garinagu (*Garifuna*)
- Ketchi Maya, Mopan Maya, Yucatecan Maya (*Mayan*)
- Mennonite (*German*)
- Mestizo and immigrants from Central America and Mexico (*Spanish*)
- Family origins in China and Taiwan (*Chinese*)

The language options on questionnaires, surveys, or commonly used forms typically present the languages previously mentioned, and "other" is the blanket option for smaller language groups interspersed throughout the Belizean diaspora. The "other" category includes students whose family origins are linked with countries in Europe, the Middle East, Africa, and India.

Pre-independence era (before 1981), Creoles were the marked majority of the cultural groups identified in Belize. Even with shifts in cultural demographics, Belizean Kriol (language) continues to be a primary lingua franca, and it is the language used on the playgrounds of many schools. Belizean Kriol is stigmatized as broken English (Decker, 1994), much like the English vernaculars or patois of the other 18 Caribbean countries with English as the official language. Subsequently, Belizean Kriol is not encouraged in formal instruction and core academic curricula. Belize's linguistic and cultural diversity issues (post colonialism, immigration, and refugee) are pragmatically overlooked given the diversity and limited resources in the church-state public education system (Decker, 1994). Nevertheless, one constant is that no matter what or how many combinations of home languages or cultures exist, all students are formally assessed only in English, the official language (Bennett, 2008; UNESCO, 2006, 2007).

Limited numbers of publications address whether Belize's demographic and school characteristic changes in recent decades might have affected the education process or achievement (school level). To compound matters, there is a blurring effect on cultural distinctions resulting from inter-cultural unions. In one study, there were clear discrepancies between what was formally presented or assigned at school and what was indicated by the same families outside of school (Haug, 1998). Furthermore, there are many communities with dynamic demographics, as evidenced by noticeable changes in proportions for ethnic groups in the last two censuses. The historic majority representation from Creole (Kriol-speaking people with British and African heritage) to the current majority being Mestizo (Spanish speaking people with Spanish and Maya heritage; including Central American immigrants) in the 2010 census reports (Statistical Institute of Belize, 2013).

In the absence of a formal measure and protocol, for all intents and purposes (valid or not), the Primary School Examinations results have been the unsanctioned or de facto indicator of School-Level Achievement. In recent years, the primary schools (*n* = 38) in one of the districts with noticeable demographic changes have not produced student-level scores to warrant being mentioned in the national top 25 list for Primary School Examinations. Trends also persisted for the comparatively lower district means for Primary School Examinations and higher ratios of teachers without specific teacher-training qualifications (Belize Ministry of Education, 2007; Belize Ministry of Education, 2012c; Policy Planning Research & Evaluation Unit, 2015). In 2013–2014, this particular district had the highest dropout rate (1.2%) and second-highest rate for grade retention (7.1% of students repeated a grade) in the country. Additionally, almost a third (31.5%) of eighth-grade students scored less than 50% on the national examinations (Policy Planning Research & Evaluation Unit, 2015).

At the time of the present study, there was no singularly recognized systematic protocol to measure school-level educational achievement in Belize. The issue of measuring school-level educational achievement was an area of interest for the Belize Ministry of Education (C. Babb, personal communication, June 20, 2018). There was an expressed need for a standardized measure for estimating school-level educational achievement and creating school profiles that are meaningful and interpretable by primary school stakeholders. By having a standardized measure, stakeholders can begin to identify and adjust relevant actions or conditions to improve achievement at the school level. Given the limited resources and differences in infrastructure across Belize, the measure has to be a low-tech, cost-effective, and efficient tool that captures relevant, multifaceted data. Replicability was also a necessity so that subsequent iterations could be used to map changes in achievement and to serve as a criterion measure to evaluate interventions.

The 2010 Belize census report operationalized educational achievement in terms of completion of primary school or higher (Statistical Institute of Belize, 2013). However, using a single-variable indicator disregards the impact of contributing factors. Furthermore, Standard 12.10 of the *Standards for Educational and Psychological Testing*, hereafter referred to as the *Standards*, discourages the use of a single source of information for high-stakes evaluative decisions in educational settings (AERA, APA, NCME, 2014). In the present study, the term “educational achievement” is conceptualized as being multifaceted and connotes recognition and attention to a spectrum of components that comprise the whole education process. For example, other studies have looked at relationships between educational achievement and various combinations of factors such as School Context, instructional process, the environment, opportunity, policies, resources, and traditions (Elliott & Bartlett, 2016; Erickson, 1987; Grantham-McGregor et al., 2007; Heafner & Fitchett, 2015; Muthén et al., 1995; O’Malley et al., 2014; Raudenbush et al., 1992; Opdenakker & Van Damme, 2007). The expectation is that this study will contribute meaningfully to the literature because of the combination of variables used to conceptualize and to estimate the multifaceted issue of school-level achievement. The school is viewed as a singular unit in which all educators contribute to the school’s level of achievement. Therefore, more emphasis is on the educators’ perspective (voice) and less on standardized testing.

In Belize, the education system can gather information to address critical issues. However, there is a need for a multipurpose product that can estimate school-level educational achievement, enable comparison of conditions within and across schools, and potentially identify exemplars or schools that could benefit from specific interventions, special programs, or targeted funding opportunities. The present study included selected educator variables, Opportunity to Learn, and school characteristics among the variables used to estimate School-Level Achievement. Two estimation protocols were compared. School-level findings were subsequently categorized and parceled as school-level educational achievement profiles. The intention was for schools and stakeholders to be able to use the product to identify comparative strengths and weaknesses regarding.

Purpose of the Study

The purpose of this study was to determine a viable method to estimate school-level educational achievement in Belize. To address the issues of the present study, the *Belize Educator Survey* instrument was developed and used to collect educator-level data on selected variables that affect the education process and outcomes. One purpose of this study was to compare estimates from two different estimation methods that drew from the same Base-Model data and determine the relative value of each model.

There is a profound need to develop school-level characterizations or profiles of educational achievement that is salient within the Belizean context. The findings from the present study could be used to create such profiles, which in turn may help schools to better understand the underpinnings of previous academic performances and inform future decisions. To increase educator support, and to garner information that was school specific, the present research study sought to integrate input from all primary school educators in the district.

Research Goal and Questions

The present study aimed to answer the following research questions given a small, unique, and finite population sample in Belize. To do so, the data collection instrument was developed and used as the primary source for educator-level information to estimate school-level educational achievement. The instrument was also intended to materialize responses that were pertinent to the development of an estimation protocol or methodology for school-level educational achievement via the following research questions:

1. Was there a difference in the overall School-Level Achievement estimates using a classic summative approach with the Weighted-Indicator Scores Protocol (WISP) versus a latent approach with a Multilevel Achievement Estimation Protocol (MAEP)?
2. Were there differences in the School-Level Achievement estimates by selected school characteristics such as School Type or Location?
3. How much variance in school-level academic outcomes was accounted for by selected variables indicating Educator Efficacy, Educators' Perceptions of Their School Environment, and the Opportunity to Learn?
4. Was there a difference between class averages by divisions (upper, middle, and lower) and school-level performance on the 2018 Primary School Examinations?
5. What was the relationship between selected school characteristic variables (School Size, Location, and Infrastructure) and variables for Educator Efficacy and Educators' Perceptions of Their School Environment?

Limitations

Belize is a member of the United Nations but has not participated in any of the major Organization for Economic Co-operation and Development (OECD) or International Association for the Evaluation of Educational Achievement (IEA) international studies. Therefore, Belize has not benefited from prior studies that operationalized Achievement Within Schools (AWS) in a national context. Modeling educational achievement is a complex issue that is further complicated when empirically tested variables used in other studies are nonexistent or exist in considerably different forms. Furthermore, the relatively sparse population is further segmented considering Belize's nine ethnic/cultural groups and associated differences in sociolinguistic registers (Statistical Institute of Belize, 2013). These issues are relevant because sample size can influence methodology decisions (based on the appropriateness of some statistical analyses).

The unit of interest for this study is at the school level. The present research study inherently covered a relatively small population. Some statistics provided by the Belize Ministry of Education were aggregated as school-level performance statistics. Notably, the present research study also addressed a specific region with a relatively sparse population. Consequently, this study was not designed for randomized sampling; all primary school educators (n = 524) of the 39 schools in the district were invited to participate.

Geographically, Belize fits in with both the Caribbean and Latin America, yet Belize is rarely sampled in studies conducted in either of the regions. Belize is one of the few Caribbean countries on the mainland, and it is rather remotely distanced from the

Greater and Lesser Antilles or the Caribbean archipelago. Though near the Latin American nations of Central America, Belize stands out as the only country in that region with English, not Spanish, as the official national language. One challenge in the present research study was to find published educational or developmental studies that included Belizean samples that could inform variable selection decisions. In terms of this study, the relative lack of published studies in Belize was further compounded by an education system that had equally limited systemic data collection in the areas of interest. The present study contributes to scholarly literature, both generally and specifically about Belize.

Significance of this Study

To have an instrument and a systematic protocol for determining School-Level Achievement allows for school-level comparisons and identification of select variables that could be targeted in school improvement interventions or programs. The outcome of the present research study can be used as a platform or segue for subsequent studies, which could further examine schools that are perceived as successful to determine if there are systematic issues or practices that set them apart and warrant replication in other schools. Similarly, schools can be identified to benefit from tailored programs or interventions.

Component factors of the present study included relevant variables that were used in other studies in varying degrees and combinations. These variables were adapted to suit the specific nuances of Belize’s education system. The intent was to create pragmatic estimates of School-Level Achievement. One potentially important outcome of this study is the quantification of relationships between the various variables used in the estimation of School-Level Achievement. Evidence to support these relationships strengthens the use of the variables as profile points and measures. Recognized variables could enable opportunities to identify exemplar qualities in schools and to illuminate factors that most likely contributed to higher achievement status. Findings about the variables could potentially inform and impact policies and action plans for sustainable improvement in school-level educational achievement in Belize.

Chapter 2

Literature Review

This chapter provides an overview of the components considered for the present research study. The first part of the chapter presents a contextual background for Belize and the term “educational achievement.” The subsequent issues covered in this chapter address selected factors that underpin educational achievement in other studies. Of central interest to this study were Opportunity to Learn; educator-related variables that included qualifications, perceptions of their school environment and climate, and Educator Efficacy; demographics and school characteristics; and quantification of academic outcomes.

Contextual Background (Belize)

The unique geography and history of Belize have afforded membership participation in both the Central American Integration System (SICA) and the Caribbean Community (CARICOM) (Statistical Institute of Belize, 2013). (See Appendix A for an index of the more commonly used acronyms in this study.) Belize is a multicultural developing nation with a population of fewer than 400,000 people and is categorized as having a low-income economy (The World Bank, 2018). Additionally, Belize is the only country in Central America in which English is the official language (Bruthiaux, 2003; Shoman, 1994; Statistical Institute of Belize, 2013).

The 1964 laws and early educational practices in Belize are variants of the British scholastic legacy to the Caribbean, based on the Westminster model (Shoman, 1994). Schools in Belize are expected to operate within the confines of the Belize constitution, protecting the fundamental rights and freedom of all students, educators, and administrators (Government of Belize, 2011). In accordance with the Belize constitution, via the Education Acts of 2003 and Amendments of 2012, schools report to district education enters and are held accountable for principles, responsibilities, and practices outlined in the *Education Rules* (Government of Belize, 2003; Belize Ministry of Education, 2012a; UNESCO, 2007) and the *Belize Teaching Service Commission Handbook* (Belize Ministry of Education, 2012b).

Religious missions in the British Honduras Settlement (now Belize) first introduced the formal public education system in Belize (Bennett, 2008). The 2010 census reported 16 religious denominations in Belize (Statistical Institute of Belize, 2013). In Belize, there are three types of schools: denominational, government, and community/private in the church-state public education system (Arcia, 2016; Bennett, 2008; UNESCO, 2007). The salaries for denominational schools in Belize are typically government funded, and the associated denominations (churches) are expected to take responsibility for management, upkeep, and spiritual guidance (Bennett, 2008; UNESCO, 2007). Specific departments and district education offices within the Belize Ministry of Education control both funding and management issues for government schools (Arcia, 2016). More than 20% of the national public expenditure for Belize goes toward education (Arcia, 2016). Although less in number, there are private schools in Belize. The costs of operation for private schools, including salaries and infrastructural upkeep, are the burden of those associated with them (parents of the students, management, and benefactors). Accordingly, private schools mostly serve children from families with the financial resources to cover tuition and fees and those students who can secure scholarships to cover the tuition and fees.

Standardization

There have been measures taken in Belize to define and standardize expectations for certain roles, responsibilities, rights, and practices in the education system with the establishment of a National Teaching Service Commission (Belize Ministry of Education, 2012b) and the Belize Education Act (Belize Ministry of Education, 2012a) that governs schools and protects the rights of educators and students, regardless of funding sources (Belize Ministry of Education, 2012a). The mission of the Teaching Service Commission is “to provide and effect a mechanism within which standards and regulations governing the quality and conditions of service of teachers will be managed with transparency, impartiality, and efficiency to ensure school effectiveness and promote public confidence in the Education System” (Belize Ministry of Education, 2012b, p. iii). This study has the potential to contribute to this area of development.

At the time of this study, there was continued interest in the standardization reform of Belize’s education system but no established protocols for measuring School-Level Achievement in primary schools. Some of the statistics reported for districts and schools include graduation rates, demographic information, and aggregated performance on the national Primary School Examinations. The scores derived from the Primary School Examinations, which is administered to eighth-grade students, play a major role in high school admissions decisions. Registration fees and other costs are relatively high for an average family, thereby increasing the stakes for score-based scholarship awards and potentially limiting access to secondary education for the poor and less fortunate. Other contributing factors that challenge access to secondary education include the lack of space in high schools and limited transportation for rural students (UNESCO, 2007).

In Belize, primary schools enroll students from grades one through eight (Arcia, 2016; Belize Ministry of Education, 2012a). As per conditions in the Convention on the Rights of the Child, in the Belize church-state education system, primary school education is relatively accessible for all children up to age 14 years and mostly tuition free for families (Arcia, 2016; UNESCO, 2007). The importance of ensuring access to schools and attendance is similarly evidenced in truancy laws for children (ages 5–14 who have not yet graduated from a primary school) and further supported with the services of school wardens and truancy officers (UNESCO, 2007). The primary function for wardens is school safety, while truancy officers address chronic absences (Humes, 2008). In a study conducted by the International Development Bank (IDB), 92% of primary-school-aged children in Belize attended school in 2009. Although admirable, that statistic was reportedly 3% lower than what was reported in the 1999 report (Näslund-Hadley et al., 2013).

With few exceptions, primary schools in Belize operate from September to June with a three-term academic year as opposed to the two-semester or four-quarter systems. In one study, Belize’s on-time completion rate (40%) for primary schools was the lowest in Latin America and the Caribbean (Näslund-Hadley et al., 2013). an indication that many students repeat one or more grade levels. Unlike in other countries, the education system in Belize does not typically implement social promotion to keep students in age cohorts or offer remediation for previous shortcomings. In Belize, a student is promoted to the next grade when they earn or exceed a prescribed minimum overall average and pass specific established subjects. Typical grade-level promotions require earning the minimum stipulated overall average of 50% to 70%, depending on the school. Some schools have additional requirements pertaining to compulsory subjects, usually mathematics and/or English. When minimum requirements are not met, students repeat the whole grade, not only the unsuccessful subject or subjects.

Promotion between the eighth and ninth grades (primary school to high school) also involves other decisions and considerations. First, some schools use individual performance on the Primary School Examinations as their primary measure for admissions decisions. Other schools use performance on the Primary School Examinations in conjunction with individual report cards (School-Based Averages) to make admissions decisions. Second, secondary school education is not free. There are fees and the cost of textbooks, materials, and uniforms to consider. The Näslund-Hadley et al. (2013) study noted that Belize had one of the lowest participation rates in secondary education in the region. In that study, there were significant differences for participation in education across ethnic groups at both the secondary and the tertiary level.

> "The Ministry of Education maintains that the exam is not intended to be the final arbiter of where students go on to attend high school, nor is it a stick to be broken over the backs of teachers and primary schools who do not perform well." (Humes, 2016, p. 1)

The preceding excerpt from the front page of a Belizean newspaper reflects the existing tension between public consumption of education statistics and education officials regarding the Primary School Examinations scores and subsequent extrapolations as a measure of school quality. Furthermore, there is implied blame for students' poor performance. The privacy laws and practices in Belize are not as stringent as those of developed countries like the USA. The Belize Ministry of Education issues an annual press release with the top 25 student performers (names, schools, scores, district), descriptive test statistics, and updates about the test, which are published by most of the local media houses (Belize Ministry of Education, 2018; *San Pedro Sun*, 2017).

It is not exceptionally difficult for interested parties to access the national or district lists of scores for individual top-performing students, including full name, school attended, and the scores earned on the Primary School Examinations. An unintended consequence of these lists is that the information is subsequently used to gauge the quality of school or district education programs without regard to scope, validity, or fairness issues. The situation is compounded because the Belize Ministry of Education disseminates education statistics that include a section that highlights top performing schools and another for the top performing students (Belize Ministry of Education, 2018, pp. 7–10). Thus, to the layperson, there is high face validity in Primary School Examinations scores for making inferences about school-level performance. Subsequent products based on the present study would provide the Belize Ministry of Education an option of a more comprehensive strategy or protocol for estimating school-level educational achievement, without the central focus being on the Primary School Examinations yet inclusive as an existing standardized measure.

Large-Scale Testing

The Belize national Primary School Examinations are intended to measure student-level overall educational achievement (Government of Belize, 2014). The design and scoring principles of the Primary School Examinations are grounded in classical test theory. The overarching latent variable assessed and represented in Primary School Examinations data is a general academic proficiency measure of eighth-grade students in

Belize (Government of Belize, 2014). Scores are used for scholarship and admissions decisions pertaining to secondary education, and they are also used to make judgments about the quality of education at primary schools. Nevertheless, the Belize Ministry of Education cautions against the use of Primary School Examinations scores as the only source of information to make high-stakes decisions such as admission to secondary school, scholarship awards, or evaluation of school and teacher quality (Government of Belize, 2014).

With heightened awareness of test-based accountability, fairness issues, and validity of score interpretations, there are psychometric reasons for not using the Primary School Examinations as sole indicator of school-level academic achievement. First, the Primary School Examinations are not equated across forms or years (Government of Belize, 2014). There are legitimate questions about differences in the levels of test difficulty across years. Second, the Primary School Examinations is only administered to eighth-grade students, the final year of primary school education. Therefore, results of the Primary School Examinations do not account for performance in other grades. School-Level Achievement estimations based on Primary School Examinations alone could mask or disregard the direct contributions of educator-level and grade-level variables in the overall school-level achievement. Third, the Belize Ministry of Education, which produces the Primary School Examinations, explicitly cautions against the sole-source use of Primary School Examinations scores for school-level or eighth-grade teacher-quality evaluation (Belize Ministry of Education, 2018). A model to estimate School-Level Achievement could help to advance a systemic culture that puts meaningful emphasis on the whole education process and reflects the involvement of the whole school (all educators and all students). Direct input from educators in the measurement of School-Level Achievement ensures that in addition to academic performance, educators' voices and experiences are actively represented in the outcome (characterization and profile).

Educational Achievement

Achievement assessments within the educational context "are designed to measure the effects of a specific program of instruction or training" (Anastasi & Urbina, 1997, p. 475). Similar to Cole's concept of educational achievement, the present study conceptualized educational achievement as a cumulative reflection of the whole education process and all of its components (Cole, 1990). Since education is a process in which the school, the family, the community, the environment, and the systemic policies contribute to levels of socio-academic educational expectations and success (Agasisti et al., 2018; Airasian & Miranda, 2002; Baker et al., 2014; Briggs & Dominique, 2011; Chaudry & Wimer, 2016; Deal & Peterson, 2011; Erickson, 1987; Evans, 2005), meaningful estimates of School-Level Achievement should also account for systemic and socio-educational factors (Masinoa & Niño-Zarazúa, 2016). The way in which educational achievement is measured and expressed conveys a sense of the goals and values of the education system (Cole, 1990).

Belize ratified and adopted the United Nations Convention on the Rights of the Child, thereby recognizing education as a right for all children (Bennett, 2008; Belize Ministry of Education, 2012a) without regard to family history. This commitment affirms the premise that all children in Belize have the right to be educated at the primary school level. However, the agreement is less explicit about the quality of education or the conditions in which education happens. Nevertheless, ratification is an indication of the value or the importance placed on access to education.

Learning and achievement are influenced by the richness and complexity of the learning environment (Burke & Sass, 2013; Chin & Chow, 2015; Kolb & Kolb, 2005; LeTendre, 2002). In other words, selected Opportunity to Learn variables such as demographics, resources (tangible and intangibles of the environment), motivation, and what others think of an individual or what they expect from the individual are variables that are likely to influence individual achievement. From an international perspective, defining or evaluating educational achievement has advanced as a result of examining cross-national studies in other areas of research in education such as sociology, anthropology, cultural psychology, and qualitative studies (LeTendre, 2002, p. 199). Furthermore, educational research in developed countries iterates the overarching premise that achievement is not solely dependent on what goes on at school (Agasisti et al., 2018; Chaudry & Wimer, 2016; Deal & Peterson, 2011; Duncan & Magnuson, 2012; Elliott & Bartlett, 2016; Farah & Hackman, 2016).

Belize is not one of the G-20 countries and has not participated in any of the international studies (Program for International Student Assessment (PISA), Progress in International Reading and Literacy Study (PIRLS), Trends in International Mathematics and Science Study (TIMSS), Programme for the International Assessment of Adult Competencies (PIAAC) used in the Stephens et al. (2015) study, hereafter referenced as the G-20 report. Additionally, the present study was intended neither as a replication nor as an adaptation of the G-20 report. However, given the scarcity of studies specific to Belize, the G-20 report was helpful for highlighting the educational achievement factors and indicators that were of interest for international studies and the G-20 nations. The studies and surveys from which the G-20 report findings were drawn are based on theories and principles that were supported by extensive research (Stephens et al., 2015). The G-20 report outlined five themes: population and school enrollment (4 indicators); academic performance (10 indicators); contexts for learning (8 indicators); expenditure for education (2 indicators); and education returns (5 indicators) (Stephens et al., 2015, pp. iii–vii).

The first theme of the G-20 report referred to population and school enrollment, which included information about primary school students, ages 5–14 years old, and high school students, ages 15–19 years old. Differences are noted in the age when children are enrolled in primary school, as well as the percentage of students enrolled in formal pre-primary educational programs. For example, 90% or more of children in the category of primary school–aged children between 2002 and 2012 were reportedly enrolled in formal education programs in all of the G-20 countries. In 2011, in the three-to-four-year-old category, France, Italy, and Germany similarly reported 90% enrollment, whereas Turkey and Indonesia reported less than 20%. Meanwhile, the USA reported 64% for the three-to-four-year-old category and achieved 90% with six-year-old children. The statistics were not intended as evidence of national interest or success but rather to show that enrollment statistics further illustrate that access in the early academic years is typically hinged upon the age-range policies for compulsory education in each country (Stephens et al., 2015). There were also differences noted in data for overall percentages or total numbers of international students by country and by age category.

The second theme of the G-20 report was academic performance. Stephens et al. (2015) sourced the more recent student data from international studies and surveys such as the 2011 administration of PIRLS and TIMSS and the 2012 administration of PISA. In this theme, 10 indicators were studied, and the report summarized the outcomes in terms of recent performance, performance on subscales in the measures (Mathematics, Reading, and Science), and changes in performance across iterations.

The third theme was the context for learning. This theme has eight indicators covering the following perspectives: students' affect for Mathematics, Reading, and Science (attitudes, motivation, and confidence), educators' context (professional development, career satisfaction, and access to specialized services), and educators' input (instructional practices and strategies). The fourth and fifth themes summarized financial aspects associated with education and the various indicators to measure and describe selected educational outcomes.

Other studies explored characteristics or issues to evaluate how the selected characteristics related to student-level outcomes (Baker et al., 2014; Deal & Peterson, 2011; Erickson, 1987; Heafner & Fitchett, 2015; Heyneman & Loxley, 1983; Schuh-Moore et al., 2012; Taguma et al., 2012). To review factors that influence achievement, the second part of this chapter discusses selected factors that have been found to be important for student learning, and this information also influenced the development of the instrument to collect educator-level data in the present study. Based on research in the field, the foci of this study included the following variables: Opportunity to Learn, School Context, Educator Efficacy, Educators Perceptions of Their School Environment, and School-Based Averages.

Opportunity to Learn

To explain the contributions of some latent variables on academic outcomes, researchers developed the concept of Opportunity to Learn. Generally, the Opportunity to Learn concept refers to inputs and processes existing within a school or unit of measure that presumably have a significant impact on achievement outcomes (Elliott & Bartlett, 2016; Muthén et al., 1995). In the *Standards*, Opportunity to Learn is presented as “the extent to which individuals have had exposure to instruction or knowledge that affords them the Opportunity to Learn the content and skills targeted by the tests” (AERA, APA, NCME, 2014, p. 56). This section of the chapter presents an overview of Opportunity to Learn as an indicator of academic achievement in terms of access and resources.

Opportunity to Learn is characterized as a multidimensional construct that is at the core of quality instruction (teaching) and a prerequisite for student achievement (Elliott & Bartlett, 2016). Opportunity to Learn variables are typically associated with processes or systems that correlate with a teacher’s effectiveness, such as instructional time, content coverage, and available resources (Elliott & Bartlett, 2016; Muthén et al., 1995). Within the general concept of the school as the unit of measure, the systemic effects embodied in the Opportunity to Learn indexes are presumed to have a holistic or incremental impact on school-level outcomes (Elliott & Bartlett, 2016). Categorically, these factors include elements of time (access to learning), content, and resources. Subsequently, school-level analyses conceivably include indexes or measures that qualify and quantify Opportunity to Learn variables in terms of time, content, and resources (Elliott & Bartlett, 2016).

Muthén et al. (1995) studied sensitivity to some Opportunity to Learn variables such as students' access to certain courses, teacher-reported variables, and outcomes on a large-scale assessment. They found that there were differences in sensitivity to Opportunity to Learn variables between Mathematics items that were content specific, problem-solving, or general. There are two noteworthy takeaway points from the Muthén et al. (1995) study. First, Opportunity to Learn issues can influence academic achievement. Second, certain kinds of items in Mathematics with higher Opportunity to Learn sensitivity (e.g., problem-solving) can potentially be used to capture movements toward adherence to curriculum and standards (Muthén et al., 1995).

Access and Opportunity to Learn

Sustainable early childhood education programs construct the foundation on which knowledge, attitudes, and values are established (Martin, 2001; Samuelsson, 2011). Access to formal education for students and the age at which they begin a formal academic career are Opportunity to Learn indicators (Stephens et al., 2015). Access to opportunities for learning also relates to instruction time and attendance rates (Elliott & Bartlett, 2016).

An Inter-American Development Bank (IDB)–sponsored study in Belize examined policies, challenges, and opportunities in the Belize educational system and indicators from 1999 to 2009 (Näslund-Hadley et al., 2013). This study sourced data from the IDB-sponsored *Labor Force Survey*, the *Living Standards Measurement Survey*, and the *Household Surveys* conducted in Belize and other countries in the region. The problem stated for initiating the study was that the Belize government had invested heavily in education and “the outcomes, in terms of access, quality, and equity, were disappointing” (Näslund-Hadley et al., 2013, p. 4).

In Latin America and the Caribbean, the gross regional mean for preschool enrollment was at 71%, while Belize reported 44.3% enrollment (Näslund-Hadley et al., 2013). Meanwhile, Belizean enrollment for primary-school-aged students in 2009 was at 93% in urban areas and 91% in rural areas. The study also looked at enrollment across student populations classified within four of the largest ethnic groups represented in Belize (Creole, Garifuna, Maya, and Mestizo). In primary school, the 2009 enrollment of Maya was 88.2%, which dropped to 40.8% for secondary school. The enrollments for Mestizo and Garifuna were at 93.8% and 96.2%, respectively, for primary school and 41.2% and 41.7%, respectively, for secondary. For Creoles, primary school and secondary school enrollments were 91.7% and 57.0% (Näslund-Hadley et al., 2013). The average primary school attainment or completion for youth 15 to 19 years of age for Latin America and the Caribbean was 88%. The national average for Belize was 83%, while the national averages for Costa Rica, Mexico, and Panama were at 95% (Näslund-Hadley et al., 2013).

In a thematic review of early childhood education centers (ECECs) in twelve countries, the Organization for Economic Co-operation and Development (OECD) reported strong correlational relationships or links between early childhood education development and sustained national growth and development (Martin, 2001, pp. 1–2; OECD, 2001). Subsequently, three major themes to promote early childhood education and care were illuminated. The first was the need for clear recognition of a nation’s role and responsibility for young children and their families. The second theme was the commitment to funding sustainable services. The third was to have or develop a cadre of adequately trained professionals to deliver engaging service for early childhood development (Martin, 2001, p. 5).

In another OECD initiative, the connections pertaining to the relationship between ECECs and national outcomes suggest that national investments in early childhood education and care are strongly correlated with health gains, the child's well-being, and the long-term academic benefits (OECD, 2009). In developed countries of the 21st century, formal education often begins before the first grade, in kindergartens, preschools, and ECECs, yet compulsory education is mostly attributed to primary school and high school (Stephens et al., 2015). Early childhood education has been recognized as an Opportunity to Learn indicator in cross-national studies, but the nexus is how a nation addresses access, availability, and quality (Baker et al., 2014; Elliott & Bartlett, 2016; Samuelsson, 2011; Schuh-Moore et al., 2012), with quality of the ECEC program being noted as a critical factor.

In developing countries (much like in developed/Western countries), cognition and socioeconomic status are strongly linked to a child's early learning in formal environments and at home (McCoy & Zuilkowski, 2015). Structural equation modeling results showed direct paths from maternal education and socioeconomic status to both verbal and nonverbal skill development in early childhood (McCoy & Zuilkowski, 2015). The results suggested that maternal education and socioeconomic factors significantly predicted language skills, nonverbal reasoning, executive functions, and more significantly, whether the child would be enrolled early or late (older than six years) in the first grade (McCoy & Zuilkowski, 2015).

Opportunity to Learn issues are also relevant for improving literacy and educational achievement in grade school. The United States Agency for International Development (USAID) sought answers to the question "How can we ensure that 100 million children learn to read by 2015" (Schuh-Moore et al., 2012, p. 1)? One initiative was to partner with the Education Quality Improvement Program (EQUIP2) "to address the core issue of the question through applied research and educational projects in multiple developing countries" (Schuh-Moore et al., 2012, p. 1). EQUIP2 developed a 12-factor index to gauge foundational Opportunity to Learn for students in primary schools.

One study looked at Opportunity to Learn as a strategy for improving educational outcomes in Guatemala, Honduras, Ethiopia, Mozambique, and Nepal. The importance of access was noted, but the study emphasized that the quality of the education delivered was a very important factor (Schuh-Moore et al., 2012). In the Guatemala case study, field research teams collected data from 10-year-old participants at 26 schools that participated in a Save the Children Program initiative. Due to the specific and unique nature of those schools, their results are limited and not generalizable to traditional public schools in Guatemala. In their study, Schuh-More et al. (2012) reported that with respect to the Opportunity to Learn factors and Reading, instruction time was reduced to 72% as a result of time lost because of late starts, early dismissals, and prolonged breaks in the day for recess. The team also noted that 63% of the students had textbooks but were observed using the textbooks only 3% of the time and other materials 11% of the time. The students' mean reading fluency was 46 words per minute, which was 14 words per minute less than the expected mean or acceptable threshold indicated for the study

(Schuh-Moore, DeStefano, & Adelman, 2012). Conclusions from the case study also indicated that the 12 Opportunity to Learn factors were not significantly correlated with learning. The process of teasing out the effects of Opportunity to Learn variables is complex, and there are times when the connections become clearer with attention on what is missing or absent. In the Guatemala study, lack of achievement and weak relationships between the study variables was attributed to the lack of instruction provided to the students (Schuh-Moore et al., 2012). Simply put, due to scheduling circumstances, the students were cheated of time dedicated to instruction and therefore deprived of the Opportunity to Learn. Time allotted for instruction is a commonly used measure to evaluate access for Opportunity to Learn (Elliott & Bartlett, 2016). The matter of time in this respect is also relevant in Belize. In Belize, earning an overall average below a school's prescribed passing mark is sufficient cause for grade retention, and a student would have to repeat the grade for another full academic year.

Content, Quality, and Opportunity

Indicators of school quality include physical infrastructure, curriculum, academic engagement, school resources, training qualifications of the adult supervisors and teachers, and outcomes on assessments (Banicky, 2000; Elliott & Bartlett, 2016). "The teacher-quality indicator most often collected is the proportion of trained teachers or those who have received the minimum organized teacher training (preservice or in-service) required by a given country" (UNESCO, 2006, p. 50). The report indicated that in Belize about half of the primary school teachers were classified as "trained." When compared to other countries in Latin America and the Caribbean, using each country's individual standards, Belize had the lowest proportion of trained primary school teachers in the region (UNESCO, 2006).

Opportunity to Learn can also be conceptualized in terms of what is done within an education system to prepare teachers, so they are able to cover content standards and meet performance expectations (Pritchard, 1995). In Belize, within the more recent decades, there have been concerted efforts to increase the level of training for teachers and implement a license requirement (Government of Belize, 2003). One such initiative is the 2010–2013 Banana Belt Primary Education Teacher Training Program that was sponsored by the European Union and delivered by the University of Belize to participating educators of the Stann Creek District (Davies & Ryan, 2013; Faber, 2010). The program aimed to address specific training needs in that region of the country and focused on primary school teachers. Education statistics for the 2013–2014 academic year show that 60.2% of teachers in the district had training in education, and 13.4% had an undergraduate degree or higher with a specific focus in education (Policy Planning Research & Evaluation Unit, 2015, p. 25).

In Belize, to become a licensed primary school teacher, one only needs to have earned an associate degree (Government of Belize, 2003). In most developed countries, primary school teachers are expected to have more training and satisfy licensing requirements. For example, the licensure requirements for elementary school teachers in the United States of America vary by state, but overall, they are more stringent than in Belize. In Nebraska, the requirement is possession of a bachelor’s degree, completed specified training, and passing scores on relevant Praxis exams (Nebraska Department of Education, 2018). Since 2014, Massachusetts requirements include: a bachelor’s degree, passing scores on the Massachusetts Test(s) for Educator Licensure (MTEL), completion of a Massachusetts Elementary 1–6 Approved Program, and the Massachusetts Sheltered English Immersion (SEI) Teacher Endorsement (Massachusetts Department of Elementary and Secondary Education, 2018).

On a broader scope, the goal of the constitution of the United Nations Educational, Scientific and Cultural Organization (UNESCO) is that nations develop a cadre of teachers who are able to and who are supported in their task of delivering quality and inclusive services to students and facilitate the attainment of universal primary education (UNESCO, 2006, 2007). There is a shortage of teachers in numerous countries across the globe, but the goal is not simply to fill classrooms with adult bodies labeled as teachers. Nevertheless, the minimum qualifications required to be a teacher in primary school (grades 1–8) vary greatly between countries and regions (UNESCO, 2006), which present challenges for cross-national comparisons.

It is not sufficient to fill educator vacancies with adults, especially if the system is lacking in its capacity to assess educator quality and engage educators in activities or programs that foster improvement. Educators matter, and “the failure of evaluation systems to provide accurate and credible information about individual teachers’ instructional performance sustains and reinforces a phenomenon called the Widget Effect” (Weisberg et al., 2009, p. 4), which implies that teachers are interchangeable, to the extent that excellence is not recognized and mediocrity is overlooked.

Taguma, Litjens, and Makowiecki, (2012) evaluated skills and staff traits associated with high-quality services and outcomes in Finland. The premise was that the quality of a teacher is important, especially for students in foundation-level grades. Their study indicated that the quality and level of training that educators had could impact eight key factors: (a) working knowledge of child development and learning; (b) quality of planned targeted lessons; (c) attending to student needs (praise, comfort, curiosity, imagination); (d) ability to elicit the children's ideas; (e) taking initiative and problem-solving; (f) capacity to build upon children's perspectives; (g) having good command of language; and (h) capacity to guide development of critical-thinking skills. Taguma et al. (2012) posited that the quality of a teacher ultimately determines the quality of instruction the teacher is able to deliver.

Policy and Expectations

Systematic policies and expectations also influence the kinds of learning opportunities that are attainable in formal academic settings. Education systems are expected to prepare each new generation for life, careers, and citizenship. With or without additional resources, education systems are expected to responsively meet and exceed a growing list of demands. In the United States, for example, with each education reform effort or trend, expectations and responsibilities of educators increase, but contact hours between educators and students remain fairly constant (O'Malley et al., 2014). Some policymakers and school administrators of the 21st century are putting more focus on links between classroom experience and achievement (Pianta & Hamre, 2009). Rather than being swept away in a tide of reform fads or being caught up in the minutia of educational politics, it may be more advisable for school districts to take stock of where they are before they decide where they would like to be and how to get there.

To date, Belize has not formally adopted a test-based accountability system within its education system. First, there are not many options in terms of large-scale standardized scores. Second, teacher evaluations are based on traditional localized supervisor observations that are reported to management, which in turn use that information to make wage increase (increments) and staffing decisions (Belize Ministry of Education, 2012a). The focus of the present research study was to estimate school-level educational achievement. However, the estimates are absolutely not intended for use as a form of school-level accountability measure or to represent a value-added measure of teacher-effect for the Primary School Examinations.

Resources and Opportunity to Learn

The matter of resources is logically linked to Opportunity to Learn. The systems of public and private education vary across and within developing countries. There are also regional differences within some countries that affect access, quality, and sustainability (Agasisti et al., 2018; Heyneman & Loxley, 1983). Another school-level indicator of Opportunity to Learn reasonably includes teacher-student ratios, contact time, and breadth of the curriculum (Elliott & Bartlett, 2016; Stephens et al., 2015). The teacher-student quotient is directly related to class size. Theoretically, smaller class sizes (increased teacher-student quotient) increase opportunity for teachers to assist individuals who may learn at a different pace or style than their peers (Raudenbush et al., 1992). Around the world, student-teacher ratios are typically affected by community size, school funding, and educational policies. The sheer numbers in more populated communities versus lesser populated communities present the issue of distribution in terms of supply and demand.

There is a meta-study of literature relevant to childhood poverty and cognitive development in Latin America (Segretin et al., 2016). That study used UNESCO’s classification; Latin America included 24 countries across Central America, South America, the Caribbean, and Mexico. Belize is officially named as a Latin American country, but of the 324 articles considered, Belize was not sampled in any of the 53 articles that met their selection criteria. Of the 24 countries in the region, samples were primarily from 13 countries. Jamaica, Nicaragua, Costa Rica, and Guatemala were each included in a single study, but El Salvador, Honduras, and Belize were not included. The review illuminated the need for sampling consideration, the importance of ecological context pertaining to characteristics of the school as a unit, and that there is a substantive gap in research publications of studies about Latin America (Segretin et al., 2016).

Educator Efficacy and Educators' Perceptions of Their School Environment

The expectations of what schools must handle go beyond academics. Efficacy of students and educators, as well as the perceptions about the school environment (school climate and school culture), impacts how both the students and educators in a school function. Although these issues are nonacademic per se, ultimately, they influence what levels of success are achieved in the school individually and collectively (O'Malley et al., 2014). "In a knowledge-based economy that makes education more important than ever, teachers matter more than ever" (Weisberg et al., 2009, p. 2).

Efficacy

"Self-efficacy is the belief in one's capabilities to organize and execute the courses of action required to manage prospective situations" (Bandura, 1995, p. 4). Self-efficacy is context specific and dynamic in terms of Bandura's (1995) conceptualization because it is one's own belief in one's capacity to accomplish (to do, or to be, or to inspire…) something specific. For that specific task or goal, one may feel confident and capable, but for another task or goal, one may feel quite the opposite. Self-efficacy may be impacted by self-love and self-confidence, but self-efficacy is not the same as either.

In a study conducted in India, Waddar and Aminabhavi (2010) studied self-efficacy and emotional intelligence in children and considered demographics across two conditions, living at home (n = 100) and living in an academic-driven hostel or boarding school (n = 100). The following quote was often used to express the core concept of self-efficacy: "If I have the belief that I can do it, I shall surely acquire the capacity to do it even if I may not have it at the beginning" (Mahatma Gandhi, as cited in Waddar & Aminabhavi, 2010, p. 340). In the summary of efficacy-related findings, the study concluded that for the children in this study, both the demographic characteristics and the context mattered. Students living at home presented higher self-efficacy scores, and the demographic variables that contributed significantly to the students living at home were gender and birth order (Waddar & Aminabhavi, 2010).

Bandura's conceptualization recognized self-efficacy as being issue specific, dynamic, and nuanced in a cycle of an individual's belief in what that person can do. Subsequently, the individual is more likely to succeed in what that individual does, particularly if that individual thinks that they can (Bandura, 1977, 1993, 1995; Bandura, Barbaranelli et al., 1996). Furthermore, Bandura theorized that if benefits are attached to success, or if there is a perception of having benefited from positive intervention, there is increased confidence in one's ability to succeed—the cycle of success begets success (Bandura, 1993). The Gandhi quote captures Bandura's conceptualization of teacher-efficacy—when educators are confident in their capacity to impart and inspire learning, the students and the community-at-large benefit from their service. Self-efficacy is a reflection of one's confidence in one's ability to wield control over personal actions (Bandura, 1993, 1995; Bandura et al., 1996), and in the context of the educator, this efficacy extends to the educator's motivation to inspire or facilitate learning and educational achievement.

During the implementation of a new History curriculum and outcome assessment, Ross (1992) studied the relationship between teacher efficacy, level of interaction between teachers and their coaches, and student outcomes. Participants included 18 teachers (36 classes) with different demographics and 6 coaches from a rural district of Ontario, Canada. The resources given to participants were threefold: curriculum materials that included detailed instructional materials, three half-day workshops through the course of the academic year, and access to designated coaches (Ross, 1992). The correlational study demonstrated positive findings on student academic achievement in the History course when teachers showed higher teacher efficacy and when teachers engaged with their coaches. However, direction and causality could not be determined (Ross, 1992). There were no significant interactions between efficacy, coaching, and achievement, which Ross (1992) suggested could also have been the result of the small sample size of the study.

Instead of keeping an emphasis on test-based accountability (external), Fullan et al. (2015) suggested that policymakers should prioritize creating conditions for internal accountability. Internal accountability is described as "a collaborative culture that combined individual responsibility, collective expectations, and corrective action" (Fullan et al., 2015, p. 4). In their model, three components sum up professional capital: individual "human capital" that includes educators and school leaders, the "social capital" in which the opportunities and atmosphere encourage teachers to learn from each other, and the "decision capital" for developing judgment and expertise over time (Fullan et al., 2015). The determination was that in a system that takes care of its professional capital with attention to human, social, and decision capitals, academic achievement is inherently nurtured and sustained (Fullan et al., 2015). To some degree, this concept was supported in Ross (1992).

Teachers' self-efficacy beliefs were measured in a study that included more than 2,000 teachers across 75 schools (junior high) in Italy (Caprara et al., 2006). The relationships between each teacher's self-efficacy beliefs and two factors, job satisfaction and students' academic achievement, were assessed. The teacher self-efficacy scale used in this study had 90 items with a seven-point response format. The measure for academic achievement was a School-Based Average rather than a standardized test. Therefore, the averages of students' final grades in two subsequent scholastic years were used. Via structural equation modeling analyses, Caprara et al. (2006) demonstrated that teachers' personal efficacy beliefs had a significant effect on their job satisfaction and on their students' academic achievement.

In another study, data were gathered from teachers in 16 urban and suburban high schools (n = 1,258 classes) in California and Michigan (Raudenbush et al., 1992). The premise of the study was that self-efficacy was contextually situated; thus, the researchers anticipated significant differences in the teacher efficacy across classes. One hypothesis was that conditions such as students' age, track, level of preparation, and class size, which varied from class to class, might predict within-teacher variation in self-efficacy. Class size was significant, but not as anticipated, because the smaller class sizes did not yield higher self-efficacy. A possible explanation noted by Raudenbush et al. (1992) was that the smaller classes were typically remedial classes for students who were low achieving. Teacher preparation was also a strong predictor of self-efficacy, although the relationship could not be explained by the engagement variables. According to Raudenbush et al. (1992), their analysis of between-teacher variation showed that teachers who were able to exercise control over key working conditions and those who worked in highly collaborative environments had higher levels of self-efficacy. Other between-teacher differences showed that women had higher levels of efficacy than males, but personal background did not have a significant effect (Raudenbush et al., 1992).

People are more likely to engage in activities to the extent that they envision success or satisfaction; “self-percepts of efficacy influence thought patterns, actions, and emotional arousal” (Bandura, 1982, p. 122). To develop a self-efficacy scale, Bandura noted that items needed to be carefully selected to reflect the appropriate construct, because perceived self-efficacy was not the same as self-esteem (Bandura, 1997, 2006). Bandura posited that the term “self-efficacy” was not interchangeable with self-esteem, locus of control, or outcome expectancies. By Bandura’s definition, “perceived efficacy is a judgment of capability; self-esteem is a judgment of self-worth” (Bandura, 2006, p. 309). Locus of control, however, is related to what one believes regarding whom (self or other) or what (nature or the system) determines a particular outcome (Bandura, 2006).

Judge and Bono (2001) conducted a meta-analysis of job satisfaction considering relationships with four factors: self-esteem, general self-efficacy, locus of control, and emotional stability. In their findings, corrected correlations between job satisfaction and the four factors showed that general self-efficacy had the most substantial positive relationship, and locus of control was the second strongest (Judge & Bono, 2001). With respect to response formats for self-efficacy scales, Bandura encouraged using scales with 0 to 100 or 0 to 10 options for participants to record the strength of their efficacy beliefs (Bandura, 2006).

Educator self-efficacy and perception of conditions in the work environment were posited as inextricably integrated into one's professional affective domain and one that influences school-climate and school-culture (Deal & Peterson, 2009). Therefore, one hypothesis of the present study was that higher levels of educators' self-efficacy and positive perceptions of conditions at the school would correlate meaningfully with academic achievement.

Engagement

In a year-long study conducted in upstate New York, researchers had a balanced sample by gender and grade of 144 elementary school students and 14 teachers (Skinner & Belmont, 1993). The study investigated time-lagged relations regarding classroom behaviors for teachers and students and the effects of perceptions of teacher behavior on student engagement. Engagement in the study referenced a motivational outcome that involved both emotional and behavioral components while carrying out learning activities. Engagement is marked by positive emotions, such as enthusiasm, optimism, effort, curiosity, and interest (Skinner & Belmont, 1993). The opposite of engagement was noted to be disaffection. Some ways in which Skinner and Belmont (1993) described disaffection in students included references to being bored, not trying hard, readily giving up in the face of a challenge, and being withdrawn.

The premise of the study was that both student and teacher classroom behavior would be mediated by their perceptions of each other's engagement. Additionally, the model tested whether reciprocal relationships regarding student engagement and teacher behavior were mediated by the teacher's perception of individual student's level of engagement. Significant correlations existed for student perceptions of the teacher behaviors and similar results were seen for teacher perceptions of student engagement (Skinner & Belmont, 1993). The overall results from the study showed indications of reciprocity. It was suggested that a teacher's involvement with each student in class impacts students' perceptions and the level of students' engagement, which in turn affects the teacher's efficacy. Students are less engaged with teachers who are less involved and who tend to be less consistent and more coercive. In summary, "strong empirical support was found in this study for reciprocal behavior between teachers' behavior and students' engagement in the classroom" (Skinner & Belmont, 1993, p. 577). Given the lack of ethnic diversity of the community, the sample was predominantly Caucasian. Therefore, the findings from the Skinner and Belmont (1993) study might be very different from findings from future or other studies if the study is conducted in classrooms with more diverse populations.

School Climate and Culture

By one Google Scholar estimation, publications by Deal and Peterson (1990, 1998, 1999, 2011, 2016) that directly address aspects of school climate and school culture have been cited more than 705, 665, 1629, 455, and 875 times, respectively. Together, Deal and Peterson have produced guides for the US Department of Education, and as such, the publications that Deal and Peterson have influenced are largely US-centric.

Nevertheless, in addition to presenting an impetus for change, or at a minimum stimulating discussions, the attention focused on their work is cited in studies that have qualified and quantified levels of impact that a school's environment (climate and culture) has had on various issues, including achievement. The guides they created for the Department of Education weigh in on day-to-day operations, educational outcomes, and reform initiatives directed for public schools in the USA.

School culture and climate are different, but they are inherently related concepts. School culture, as described by Deal and Peterson, is the collective web of formal, informal, overt, and covert "underground stream of norms, values, beliefs, traditions, and rituals that have built up over time as people work together, solve problems, and confront challenges" (Deal & Peterson, 1998, p. 28). School culture is ingrained in the identity and operation of a school because it is developed over time. However, even the most established traditions at the core of school culture can be changed (destroyed or improved) with the right conditions (Deal & Peterson, 2011). School climate, though, is temporal, in-the-moment, and school climate is easily affected by current events (Deal & Peterson, 1999, 2009, 2011, 2016).

School climate is a reference to the conditions existing at a point in time, like a snapshot or picture. There could be a sense of joy and celebration permeating the school shortly after the school wins a competition. In the same school, the atmosphere could become quite morose if there is news of tragedy halfway through that same day of the victory. School culture, however, is more like a full-length movie featuring traditions, expectations, values, relationships, communication, and purpose. School climate and culture are projected in the perceptions of educators and manifested in relationships and school practices. For these reasons, school climate and culture were of interest in the present study, in terms of the educators' perceptions of their school's environment.

In another study, the relationship between school climate, family structure, and academic achievement was investigated (O'Malley et al., 2014). The study was based on data collected from public high schools (*n* = 902) in California in grades 9 and 11 from 2008 to 2010. There were approximately 1.5 million students, but the study conducted by O'Malley et al. (2014) was limited to schools that had administered the *California Healthy Kids Survey*. The study sample had 305,956 (61.6%) students in 9th grade and 190,946 (38.4%) in 11th grade (O'Malley et al., 2014). The first research question was "Are students' positive school climate perceptions associated with improved academic outcomes for youth living in different family structures (i.e., single-parent, two-parent, foster care, and homeless)?" The second research question was "Are students' positive school climate perceptions associated with a reduction in the academic achievement gap commonly observed between at-risk youth and their peers?" (O'Malley et al., 2014, p. 4).

In the *California Healthy Kids Survey*, family structure was measured by a single item that asked students where they lived. To respond, there were 12 options, and students were instructed to select the option with the description that best matched their situation. In O'Malley et al. (2014), family structure was further collapsed into four categories. Of the students in the sample, 26.6% (*n* ~ 132,175) lived in single-parent homes, 69.6% (*n* ~ 345,843) lived in two-parent homes, 1.8% (*n* = 8,582) were homeless, and 0.5% (*n* = 2,310) lived in foster care homes (O'Malley et al., 2014). Similarly, the academic outcome indicator was measured using a single item, a self-reported grade-point average (GPA) that described the student's perception of their grades during the past year.

School climate was measured using a five-point Likert scale response format, similar to the one developed by Hanson for the *California Healthy Kids Survey* (Hanson, 2012). The four constructs covered in the school climate indicator were relationships with adults at the school (six items), opportunities for meaningful participation in school (three items), perceived school safety (two items), and school connectedness (four items). The internal reliability for the constructs ranged from coefficient alpha of .69 to .88 (O'Malley et al., 2014).

The relationships between students' perceptions of school climate, their family structure, and self-reported GPA were estimated with regression analyses. "The regression models were estimated using a multilevel, random-intercept approach that accounted for the clustered nature of the data (i.e., students in schools)" (O'Malley et al., 2014, p. 7). School-level analysis was not the intended goal of their study. The researchers noted that they opted for the multilevel approach because the parameter standard error estimates are more conservative in a multilevel approach that accounts for common variance in the outcome variable among same-school students (O'Malley et al., 2014).

Findings from the O'Malley et al. (2014) study indicated that family structure had a differential effect on the academic outcome: Single-parent, two-parent, homeless, and foster care had average GPAs of 2.96, 2.72, 2.60, and 2.59, respectively. However, across all the family structures "higher self-reported GPA was associated with more positive school climate perceptions" (O'Malley et al., 2014, p. 9). Positive school climate perceptions supported increased chances for academic achievement across family structures, especially for youth living in challenging circumstances based on the positive moderating effect of perceptions about school climate on academic outcome. For example, O'Malley et al. (2014) explained that students with average perception of their school climate tended to have a grade expectation of B, which was similar to the expectation of students in both the homeless situation and the two-parent homes. However, when the student had very low perceptions of the school climate, the student in the homeless situation would expect a C, but the student in the two-parent home would expect a C+, a half grade higher (O'Malley et al., 2014). This study was relevant to the present study because of the nature of the indicators, the methodology, and the findings. Strong correlations between positive perceptions of school climate and academic outcome corroborates inclusion of components of both variables in the process of estimating school-level achievement.

School Characteristics and School Context

The term "context" as used in this study references the characteristics of a school's environment, as the sum of its demographics (students and educators), systemic nature (location, management, teacher-student ratio), and infrastructure. The abstract from a German paper brought attention to the Heyneman-Loxley effect (Zumbach, 2010). The Heyneman-Loxley effect posits that to determine School-Level Achievement in developing countries, school characteristics were the better indicator of achievement than family socioeconomic status (Heyneman & Loxley, 1983; Zumbach, 2010).

Physical conditions affect human interactions in one way or another. Infrastructural inadequacies in schools can also present safety issues, impede interactions and mobility, hinder access to play, reduce variation in learning spaces, and put additional constraints on pedagogy. In a study sponsored by the World Bank, researchers examined the effect of school and teacher quality on academic achievement in primary schools across 29 high and low-income countries. Indicators of school quality included, but were not limited to, infrastructural facilities, admissions policies, academic and nonacademic Opportunity to Learn elements, school financing, and school assets (Heyneman & Loxley, 1983). Examples of infrastructural elements were the size of the campus, type, age, condition of structures, and whether or not the primary school had a school library (presence, size, number of books, usage), an auditorium, a science laboratory, a playground, or dedicated space for sports.

The Heyneman and Loxley (1983) study also posited that both the quality of the school and the quality of teachers appeared to have great influence on student learning. The Belize Ministry of Education recognizes that the qualifications of educators are indicative of the formal training received specifically for the education profession and by extension level of trained educators at a school. In recent decades, attention was brought to the proportion of trained teachers in the southern districts of Belize, which at one time was much lower than in other parts of the country (Belize Ministry of Education, 2007). One of the many ways in which attempts were made to change the situation was an intervention project dubbed the EU/UB Banana Belt Project. The initiative was sponsored by the European Union (EU) and delivered by the University of Belize (UB) to teachers and school administrators (principals) of the region (Faber, 2010). One aspect was addressing specific themes and issues geared toward primary school principals, managers in the Banana Belt (Davies & Ryan, 2013). The training sessions were also conducted locally in their region to reduce the logistics associated with inter-district travel and minimize disruption to school operations. Such efforts draw attention to the importance of helping educators to help themselves to be better educators and improve school quality in developing nations. As indicated by Heyneman and Loxley (1983, p.1184) “the poorer the national setting in economic terms, the more powerful this school effect appears to be,” so School Context matters and should be included in conversations and studies about educational achievement.

Demographics

The diversification of a population adds to cultural richness of a nation but not without some growing pains or unintended consequences. The relevance of immigration issues to this particular study is the ethnic plurality, language diversity, and politicization in the aftermath of relatively quick demographic changes in Belize. Belize’s education system was in its infancy with the first major era of demographic changes. However, as a developing nation in the 21st century, the challenge is profound for addressing the needs of students of families from Central America who are seeking refuge in other countries as they flee the atrocities of their home countries.

Long before Belize became an independent nation, it was a small, multicultural colony of Great Britain. In the nineteenth century, the population demographics of Belize (then known as British Honduras) changed with the inclusion of people from several countries. In 1802, the Garinagu people who spoke the Garifuna language came from Honduras fleeing civil unrest and settled in the southern part of the country (Shoman, 1994). The Ketchi Mayans left Guatemala to find peace and make their home in their ancestral lands in southern Belize (Bolland, 1997) with other Ketchi of British Honduras.

The Yucatecan Mayas settled in northern Belize, following the 1847 Guerra de Castas (Caste Wars) in the southern provinces of Mexico (Shoman, 1994). Post-emancipation of slavery, between 1844 and 1917, there were the East Indians and Chinese who were brought in as indentured laborers to work in the British territories of the Caribbean, including British Honduras (Shoman, 1994). In the 1950s, the Mennonites were also given permission to establish several communities that were distinguished from other Belizean communities because of their religion and language (Shoman, 1994).

Post-independence (1981), Belize became home to refugees fleeing civil unrest in the neighboring Central American countries of Honduras, El Salvador, and Guatemala. The number of Latin American or Spanish-speaking people who came to Belize, combined with the existing Mestizos, who also speak Spanish, have increased to the point of becoming the linguistic majority group within Belize. In effect, the inclusion of the Central American refugees significantly changed the demographic landscape of Belize (Bolland, 1997; Statistical Institute of Belize, 2013).

Ethnic heterogeneity has wide-ranging implications in studies within the behavioral sciences (McKibben, 2004). In 1980, the year before Belize’s independence from Great Britain, the population of Belize was listed as 145,353 people. In the national demographic distribution, Creoles made up 40.0% of the population, and Mestizos (including Central American refugees) were 33.4% (Statistical Institute of Belize, 2013). The 1991 census report indicated a population increase to 189,392 and a shift in the demographic distribution to 29.8% Creoles and 43.6% Mestizos. This trend continued and is evidenced in the 2000 census with the Creole and Mestizo ethnic groups accounting for 24.9% and 48.7% of the 240,204 people in Belize, respectively (Statistical Institute of Belize, 2013).

In the 1980s and 1990s, civil unrest and socioeconomic disparity occurred in El Salvador, Guatemala, and Honduras, so many of their citizens sought refuge in Belize. The population demographics and sociocultural dynamics changes are evident in all districts, but the changes were most dramatic in the southern districts, where agricultural work was available in the citrus industry and the banana belt (Statistical Institute of Belize, 2013). Based on the 1980–2010 census reports, between 1980 and 1991, in the southern part of Belize there was a remarkable population growth, particularly in the rural areas of the Stann Creek District. There, the census statistics showed a 54.92% increase in the population (Statistical Institute of Belize, 2013). By the 2010 census, the population of Belize had increased to 324,528 persons, and the national demographic distribution had flipped between the two larger groups to 25.9% Creole and 52.9% Mestizo. The distribution specific to the Stann Creek District reflected ".9% Asian, 1.7% Caucasian, 22% Creole, 5% East Indian, 27.5% Garifuna, 16.9% Maya, .2% Mennonite, 33.9% Mestizo, and 1.2% other and not stated" (Statistical Institute of Belize, 2013, p. 20). These statistics bear relevance to sociocultural dynamics and educational practices and measures within a small country like Belize.

With respect to the most recently published census demographics, Belize has approximately 350,000 people, and the population can be classified into nine distinct ethnic/cultural groups and "other" for those who identify differently or choose not to affiliate (Statistical Institute of Belize, 2013). This statistic is important because it references significant cultural differences and indirectly highlights the need for that aspect to be covered in teacher training programs in the region. Erickson (1987) has posited that people tend to make sense of their complex world by categorizing people (stereotyping) in ways that may or may not hold true, but nevertheless, those stereotypes influence subsequent actions and interactions. Given the general ethnic multiplicity in Belizean classrooms, and the existence of pockets of classes with single-culture representation, educators in Belize could benefit from sensitivity training to teach in both single-culture and multicultural classroom environments.

In general, the expectation from teacher training programs is that participation in a multicultural education course is most likely to have a positive impact in reducing stereotyping attitudes that can be counter-productive in educational settings (Tran et al., 1994). To that end, diversity (students, educators, and curriculum) has also been linked positively to better quality teaching and learning achieved (Hurtado, 2001; Tran et al., 1994). Ethnic diversity could be a significant factor if the findings in other countries hold true for Belize, thus it is one of the School Context variables explored in this present research study.

Making inferences about educational outcomes warrants the combination of descriptive data with other data that contribute to the argument or theoretical relationship at hand (Smith, 2002). Cross-sectional studies showed strong relationships with socioeconomic status and school readiness skills in India, Ecuador, and Jamaica, while representative national surveys in 16 Latin American countries also had similar findings for socioeconomic status as a predictor of completing high school (Grantham-McGregor et al., 2007). Another assertion is that cultural context impacts educational outcomes (Purves, 1987). Based on different studies conducted in numerous countries, the

International Association for the Evaluation of Educational Achievement (IEA) acknowledges that education systems have characteristics that are unique to their particular contexts, and as such, researchers must consider both the cultural and educational history of the education system they study (Purves, 1987).

The notion of culturally driven characteristics is also supported by those who evaluate teachers and schools. What constitutes a “good teacher” or “good school” is not a universal concept (certainly not a national concept in diverse nations and communities). What makes the teacher or the school “good” is dependent on or grounded in the tapestry of the expectations, exposure, cultural influences, and the measures they apply to validate learning (LeTendre, 2002). Therefore, in the midst of diversity, instead of creating additional extensive lists of variables, educational researchers are encouraged to use or emphasize participant perceptions of their school environment, individual-level interactions, and cultural norms (LeTendre, 2002; Plomp, 1990).

Language

In Belize, there is linguistic and cultural diversity owing to a history with three major groups of Maya (Mopan, Ketchi, and Yucatecan); British colonialism that involved acquisition of resources, spreading of Christianity, African slavery, and Asian indentured servitude (Indian and Chinese); and subsequent immigration of Mennonites, Lebanese, and refugees from Central America (Bolland, 1997; Shoman, 1994). Cultural plurality, low population density, and limited fiscal resources have necessitated and fostered an education system that assesses students only in the official language, English, which is not the first language of the majority of students (Bennett, 2008). Most students speak a language other than English at home and with their peers. The process of language acquisition and acculturation is significantly different for first-language acquisition than it is for second-language acquisition (formal or informal) after home language(s) and culture are acquired and stable (Cummins, 2016, 2017). The challenge associated with learning a new academic language is further compounded when the language learner is a child of an immigrant who also has to contend with complications associated with differences in educational systems in addition to social factors that brought them to the new language environment (Cummins, 2016).

As an educator in Belize, this author had the observation that more care was needed for textbook selection (especially for Reading as a subject) because many textbooks and literature available to educators in Belize were written in developed countries. Naturally, those textbooks and literature were grounded in cultures that were foreign to Belizean students. Another issue was that many students experienced language interference, most likely attributable to a lack of differentiation between academic English and the lingua franca, which is an English-based Creole (Kriol). Linguistic and cultural diversity is relevant to the present research study since there is a gap in the literature for this issue with the Belizean context. Locally, there are anecdotal differences between rural and urban situations that range from culturally diverse communities with technology and modern conveniences to remote monoculture communities with limited or no telecommunication infrastructure and minimal exposure to English outside of school. Additional research is needed to empirically substantiate a position regarding language and academic achievement in the culturally diverse communities of Belize.

Socioeconomic Status

In numerous academic achievement research studies and publications, strong correlations have persisted between culture, socioeconomic status (community and family), and achievement variables (Anderson et al., 2007; Duncan & Magnuson, 2012; Erickson, 1987; Evans, 2005; Heafner & Fitchett, 2015; Segretin et al., 2016; Sharkins et al., 2017). However, it is important to note that having a low socioeconomic status label is not always equivalent to low academic achievement. Lack of financial resources does not automatically equate to academic failure because there may be other factors, such as hope and resilience that can mediate success (Agasisti et al., 2018; Dixson et al., 2018). Resilience refers to one's capacity to "prosper despite encountering adverse circumstances" by way of skills that are essential to play active roles in communities and for life-long learning in this century (Agasisti et al., 2018, p. 4).

Socioeconomic status and poverty constructs in many studies are products of multiple single indicators such as family income, parental occupations, percentage of students on free and reduced lunch, mean maternal education level, and per-student mean of federal subsidy or operation costs (Chaudry & Wimer, 2016; Duncan & Magnuson, 2012; Segretin et al., 2016). Studies have related poverty to a number of negative outcomes (Chaudry & Wimer, 2016; Duncan & Magnuson, 2012; Farah & Hackman, 2016; Segretin et al., 2016; Sharkins et al., 2017). A strong correlation between indicators and outcomes over several studies does not singularly justify causality, nor does it negate examination of other contextual variables.

Context is important because socioeconomic status and poverty are as unstable as they are complex; incomes change, natural disasters happen, opportunities manifest, and people adapt (Duncan & Magnuson, 2012). As noted in Duncan and Magnuson (2012), some people experience a brief period or periods of poverty. For others, it is an intermittent situation. Then there are those for whom poverty is long-term. The issue of socioeconomic status is generally considered as pertinent in both health and education fields, and there are studies in which the findings indicate that both the timing or duration and the level of poverty status can directly impact a child's cognition, physical development, and health (Chaudry & Wimer, 2016; Farah & Hackman, 2016; Valadez, 2010). Although socioeconomic status is often acknowledged as having an impact on achievement, the construct should not be treated as a simple straightforward variable to include in research designs without a thorough study of the cultural and geographical context and operationalization (Chaudry & Wimer, 2016).

Educational Investment

Congruent with human capital theories, international funding agencies invested in education policy, infrastructure, and programs within developing countries (Martin, 2001). Operational costs (salaries, infrastructure, maintenance, programming, etc.) typically represent a government's financial investment in education in developing countries. However, in developing countries where the fees are low or there is no fee, parents are often significantly burdened with other peripheral costs associated with education (books, uniforms, fees, transportation, etc.) (Akaguri, 2013; Härmä, 2016). For the poor, peripheral expenses a create barriers to access even for those who are awarded scholarships (Akaguri, 2013; Härmä, 2016; Tooley & Dixon, 2006).

In order to meet commitments outlined in international agreements regarding education for all, Akaguri (2013) suggested that Ghana, and other countries with similar circumstances, develop and implement more definitive pro-poor policies that remove associated educational costs from the poorest households. Härmä (2013, 2016) has drawn attention to issues that influence enrollment in Nigeria, such as the availability, quality, and differences in the realities of urban and rural settings. With reference to Ghana, India, and Nigeria, despite poverty, low-fee schools and semiprivate schools (nongovernmental) were on the rise. The increase of nongovernmental schools was proportional to public opinion that the government schools were being outperformed in Mathematics and English, even though the government expenditure for salaries was higher for the government schools (Tooley & Dixon, 2006). At the micro-level, whether students have the essential materials for schools is a reflection of investment in the students' education, regardless of who sponsors the materials (Akaguri, 2013). The issue of investment in education can also be tied back to the Heyneman-Loxley effect, which iterates that socioeconomic status is not the best predictor of academic success in developing countries; the overall school condition is a better predictor of achievement (Heyneman & Loxley, 1983; Zumbach, 2010).

School Types

In a study conducted in secondary schools of Flanders (the Dutch-speaking part of Belgium), Opdenakker and Van Damme (2007) examined the relationships and effects between school outcomes and school characteristics. The sample in Opdenakker and Van Damme (2007) included 57 mainstream high schools. In Flanders, there are more private denominational (Catholic) schools (68.4%) than public schools (31.6%), and both types of schools received some level of government funding (Opdenakker & Van Damme, 2007). Public schools all had mixed-sex (co-ed) student populations, but the Catholic schools varied; some were single-sex and others were mixed-sex.

In terms of academic achievement outcomes, the study only looked at Mathematics. In their model, Opdenakker and Van Damme (2007) conceptualized school characteristics in three parts: composition of schools (student population, teaching team, and school leader), school practices (educational framework, organization and management, work and learning environment, and class climate), and School Context (denomination, school size, curricula, and facilities). Student-level indicators were based on family and personal characteristics. Class-level indicators were overall student characteristics, teacher characteristics, and classroom practice.

A key finding was that in the between-school analysis, 89% of the variance in achievement in Mathematics and 56% of the variance in effort could be explained by the variables in the model. Another notable finding was that school size mattered as it contributed 31% of the explained variance in teacher relationships or cooperation at the school. Furthermore, educator consultation and cooperation had a strong positive effect on school-climate characteristics. This strong positive effect is also associated with favorable academic achievement outcomes (Opdenakker & Van Damme, 2007). It was also noted that student composition mattered significantly with respect to relationships between educators at the school and the learning climate. Finally, school type mattered in their study. The authors indicated that their findings confirmed results of other studies regarding the positive effects on achievement that have been associated with attending Catholic schools (Opdenakker & Van Damme, 2007).

The approach taken in that study recognized the interconnectedness of variables in an education system and took into account possible relationships and influences between school characteristics (Opdenakker & Van Damme, 2007). Additionally, Opdenakker and Van Damme (2007) noted they found in previous experience studying school effectiveness in the region that in order to investigate the effects of the selected variables, they also had to pay attention to interactions between variables and mediator effects. The Opdenakker and Van Damme (2007) study was conducted in a school system where most schools in focus were Catholic. However, findings in that study are not necessarily generalizable to Catholic schools in the present study, given that there are uninvestigated differences between Flanders and Belize.

Assessment and Measurement

Should school-level educational achievement be dominated by achievement scores from large-scale assessments? Many education systems struggle to create accountability measures. Systems justify reliance on test scores in accountability efforts for reasons such as: cost factors, validity, fairness, representativeness of the educational process, or sensitivity to differences in teacher effectiveness (Airasian & Miranda, 2002; Weisberg et al., 2009).

The age-old practice of assigning grades and computing overall averages is recognized as an inherent task for educators across majority of school systems. Typically, educators combine marks or averages from multiple subject-level grades to create an average that represent student's overall achievement. In most education systems, including Belize's, schools use student averages to determine overall class-level performance. This process is often further extended to use class-level performance to determine school-level performance. However, there are issues associated with school to school comparisons using this age-old practice that are worthy of consideration. For example, there are school-to-school differences in student learning opportunities, differences in curricula, in difficulty levels of assessments, and in the grading schemes used to compute grades and grade-point averages, and differences in teacher quality between and within schools (Weisberg et al., 2009). Furthermore, the quality school-based assessments and student performance on said assessments resonate on the effectiveness of the educators in a school.

The decision to rely on students' test scores to pass judgment or make inferences about the quality of instruction at any given school involves many assumptions. One of the biggest presumptions about large-scale assessments is that a test is sensitive to instructional practices (Airasian & Madaus, 1983; Airasian & Miranda, 2002; Amrein & Berliner, 2002; Moss et al., 2005). No matter how many times that presumption was warranted, consideration of other factors (test validation) is prudent before making decisions or designations of accountability based on students' scores. Test scores are valid for specific populations, purpose, and point in time. Therefore, not all tests should be regarded as a reflection of instructional content, practices, or quality, whether positively or negatively.

School-Based Averages as an outcome of academic performance include results from teacher-made curriculum-based assessments (formative and summative) that evaluate progress toward acquisition of instruction or curricular material. One study investigated the relationship between curriculum-based measurement and standardized assessments (statewide assessments and proprietary measures) using stratified samples, based on socioeconomic status, taken from two school districts in Pennsylvania (Shapiro et al., 2006). Two subject areas, Mathematics and Reading, were the foci of the study. Data were collected on three occasions for curriculum-based assessments during a normative study conducted prior to data collection on the various standardized measures used in the comparative study. The findings in the study indicated that across school districts there were moderate to strong correlations between the curriculum-based assessments in Reading and Mathematics and the various standardized tests (Shapiro et al., 2006). Given the data and results from the study, the implication is that curriculum-based measures are a likely source of information when projecting students' chances for success on standardized measures. At the high school level, there are studies that indicate that GPA is a very strong predictor of academic success at the college level (Geiser & Santelices, 2007; Hodara & Cox, 2016). The notion is that the high school GPA is "a measure of cumulative performance over time and thus quantifies other skills or competencies—beyond reading and math proficiency—that are necessary to succeed in college" (Hodara & Cox, 2016, p. 12). Standardized tests are traditionally presented as a more uniform and valid yardstick for assessing student ability and educational achievement than high school GPA. Geiser and Santelices (2007) challenged the position that high school GPA was not a reliable criterion because of differences in grading standards across high schools. The result of their study showed that high school GPA was a much better predictor of freshman year and long-term academic performance (Geiser & Santelices, 2007, p. 24).

Correlational analyses of student scores from large-scale measures to demographic variables such as socioeconomic status, ethnicity, and language differences have highlighted the issue of an achievement gap across the USA (Anderson et al., 2007). The achievement gap can be explained as the difference in academic performance on large-scale assessments observed between the higher-achieving and lower-achieving groups of students. Higher achievements are typically observed with White and Asian students, while lower achievements are typically observed in African American, Hispanic, Native American, and Pacific Islander student groups. Acknowledging and addressing the achievement gap has become a pertinent matter for the education system, the economy, the social stability, and the moral health of the nation (Evans, 2005). Conceptualizing and articulating the nuances of the achievement gap subsequently influences how researchers measure school achievement and what we can be done with relevant knowledge (Anderson et al., 2007).

"Consideration of measurement issues for any assessment should start with the identification of the purpose of the assessment" (Linn, 2002, p. 28). The use of test statistics developed for one purpose and used for another, such as for the expressed purpose of determining school-level accountability, is a practice that threads a fine line for appropriate use. With respect to achievement gap studies, Evans (2005) first posited that "the origin of the gap lies neither in the students nor in the schools" and, second, that variables such as "skin color, ethnic status, and poverty by themselves, do not determine a student's performance" (Evans, 2005, p. 583). On the first notion, the strong relationship between socioeconomic factors and lower achievement is recognized, but at the same time, it is also acknowledged that neither the school nor the student is in control of those factors. The second notion acknowledges that socioeconomic factors linked with low academic achievement are not causative factors as there have been cases of students with low socioeconomic status having high achievements (Agasisti et al., 2018; Evans, 2005; OECD, 2011).

The *International Test Commission's Guidelines on Test Use* (hereafter referred to as the *ITC Guidelines-TU*) were designed recognizing differences across countries. The benchmarks of the *ITC Guidelines-TU* are such that "they can be used to compare or align local standards to establish international consistency and comparability" (International Test Commission, 2001, p. 7). The fundamental principles outlined in the first part of the *Standards for Educational and Psychological Testing* (the *Standards*) apply to all stages of test development and use, including evaluative measures applied at the institutional level (AERA, APA, NCME, 2014). Both the *Standards* and the *ITC Guidelines-TU* encourage attention to psychometric properties of assessment in terms of validity, reliability, and fairness. The *Standards* define validity in terms of "the degree to which evidence and theory support the interpretations of the test scores for the proposed uses of the tests" (AERA, APA, NCME, 2014, p. 11). Based on the *Standards*, validating an evaluative measure should be addressed in terms of establishing the intended uses and interpretations, information about the samples and validation processes, and presenting evidence of validity (AERA, APA, NCME, 2014).

Test scores, standardized or not, represent the quantification of an examinees' behavior in response to an item or some items that sample a specified domain/ or content area(s) at a given point in time. Within the public School Context in the USA, whether explicitly or implicitly, scores from standardized tests have generally been accepted as indicators of district-level and school-level quality (Erickson, 1987; Fitchett, 2015; Haladyna et al., 1991; Heafner & Hodge & Welch, 2016; Holbein & Ladd, 2017; Murphy, 2012; Paufler & Amrein-Beardsley, 2014).

Standardized testing is a widely accepted tool that can be used to assess students across subgroups with the same content and testing conditions. The standardization of the assessment process is essential for fairness, but standardization does not necessarily ensure fairness (Geisinger, 2000). Furthermore, as often indicated by test publishers, a standardized test in and of itself can never represent the complete mapping of school achievement (Geisinger, 2000; Haladyna et al., 1991).

The use of test scores in school accountability measures meets resistance especially when those who are mandated to implement the testing were not included in the development or legislative process (Hodge & Welch, 2016). To preempt resistance, the present study engaged stakeholders in processes leading up to the estimation of school-level educational achievement in Belize. The input from educators and the Belize Ministry of Education were actively sought and incorporated. The education system in Belize does not currently administer large-scale standardized testing to the extent as is done in developed countries or other test-based accountability systems of education. Instead, the conceptual framework in this study considers a whole school approach that acknowledges the multifaceted nature of education and recognizes that many factors contribute to any definition of success, including School-Based Averages as one of many outcomes.

Notable Frameworks

The educational achievement construct is complex, so the process of determining which variables, factors, and indicators to include in a location-specific model is neither simple nor straightforward. Grounded in years of research, the following three frameworks exemplify the mindset of a holistic approach to evaluation in education:

Charlotte Danielson's Framework for Teaching Evaluation, Robert Marzano's Teacher Evaluation Model, and the Institute of Education Sciences' G-20 study. Essentially, the frameworks each operationalize four similar domains critical to the education process and evaluation of educators: planning and preparation; the classroom environment; instructional strategies; and collegiality and professional responsibilities (Danielson, 2007; Marzano, 2013).

Frameworks inherently provide a means to increase awareness, facilitate improvement, and promote measures to improve educator effectiveness. The overall contribution attributed to the educators' input in the education process is tantamount to the educators' effect on achievement. There are two key differences between conceptualizations of School-Level Achievement used in previous studies or frameworks and the present study. Other studies used standardized test scores in their process to determine achievement and outcomes are tantamount to the quality of the educator's input or effectiveness. In this study, the two methods of estimating School-Level Achievement are compared, and both include averages from classroom assessments and the educators' voices (Educator Efficacy and Educators' Perceptions of Their School Environment). Educators' input is included as a pivotal factor in estimating achievement (forward) as opposed to achievement scores being the indicator of educators' input (afterward).

Based on previous country-level educational assessments, the G-20 report covered the following indicators: Opportunity To Learn, Teacher Training, School Type, Socioeconomic Status, Population Demographics, School Environment, and school Climate (Stephens et al., 2015). It is anticipated that there are differences in contextual factors within management systems and those differences may introduce variables with the potential to influence achievement. The problem is that the context and characteristics of Belize are markedly different from education systems that use the Danielson or Marzano frameworks and from the countries studied for the G-20 report. Additionally, educators and the education processes are governed by the Belize Education Act of 2010, and the act is based on legal positions, not educational research or learning theories.

Protocols

In reference to the two methods of estimating School-Level Achievement in this study, the Weighted-Indicator Scores Protocol (WISP) and the Multilevel Achievement Estimation Protocol (MAEP), considerations were given to the small, unique population sample in Belize and the limitations therein. A single culturally sensitive data collection instrument was developed for this purpose. Also guided by issues and options discussed in previous studies, this study sought to answer five questions. The first question pertained to the two methods. *Is there a difference in two methods of estimating school-level educational achievement?* One method was a classic summative approach, the WISP. The other method used the latent approach, the MAEP.

The types of analyses used in the WISP were inarguably classic; they were established in the era before the first version of the internet came into effect in 1969 (Gregersen, 2019). Factor Analyses (FA), Eigenvalues, Multivariate Analyses, Pearson Correlations (*r*), and the value of intelligence tests were topics of discussions and publications in the 1920s. In the 1950s, different facets of reliability were on the forefront when the formula for coefficient alpha (internal consistency) was published in the early 1950s (Cronbach, 1951). The primary classic analyses used for the WISP were Reliability (internal consistency), Analysis of Variance (ANOVA), and Confirmatory Factor Analyses (CFA). On the other hand, the MAEP used a more contemporary iteration of multilevel structural equation modeling (MSEM) that was published in the 21st century (Muthén & Muthén, 1998–2017). The novel multidisciplinary approach gave rise to Structural Equation Modeling (SEM) in the 1970s and advanced more complex methodologies such as growth models, latent class growth models, generalized linear models, and MSEM (Matsueda, 2011). In general, multilevel analyses take into account the nested nature of samples, such as students in classes are in a school that is a part of a community or a school district.

Belize is a developing country, and the issues faced are not always aligned to those of developed countries. For example, the Heyneman-Loxley effect indicated that in developing countries, school characteristics were a better predictor of School-Level Achievement than socioeconomic status (Heyneman & Loxley, 1983; Zumbach, 2010). Therefore, the second question in this study pertained to some specific school characteristics. *Were there differences in the School-Level Achievement estimates by selected school characteristics such as School Type (Catholic, Non-parochial, or Other Denominations) or Location (Urban, Rural, or Remote)?* In addition to the school characteristics and type, a central force in every school is the educator. Logically, the educators' voice needed to be included in the estimation process. A portion of the survey instrument measured Educator Efficacy and Educators' Perceptions of Their School Environment. Subsequently, the third question in this study was included. *How much variance in school-level academic outcomes was accounted for by selected variables* *indicating Educator-Efficacy, Educators' Perceptions of Their School Environment, and the Opportunity to Learn?*

One of the frequently asked questions in educational research seeks to identify what factors predict future academic performance. In the absence of relevant grade-level standardized measures in Belize, other variables must be considered in models for estimation of School-Level Achievement grade-level performance and viability of teacher-made assessments. The fourth question in this study stems from consideration of whether grade-level school-based assessment outcomes are good predictors of external assessments. *Was there a difference between class averages for upper, middle, and lower divisions and school-level performance on the 2018 Primary School Examinations?*

The fifth question in this study once again puts attention on the educator as an important element in the education process in consideration of the Heyneman & Loxley effect and the self-efficacy theory. Instead of focusing on socioeconomic status as a predictor of achievement, the Heyneman & Loxley effect highlights look at school characteristics as a better predictor of achievement in developing counties. Meanwhile, the central premise of Educator Efficacy is the connection between what educators think they are capable of and what they proceed to accomplish (Bandura, 1995). In summary, the question asks: *What was the relationship between selected school characteristic variables (School Size, Location, and Infrastructure) and variables for Educator Efficacy and Educators' Perceptions of Their School Environment?*

Chapter 3

Methods

The *Belize Educator Survey* Development

The primary data collection instrument used in this study was the *Belize Educator Survey*, which was designed for this study. This instrument was piloted in June 2018 with the approval of the University of Nebraska Institutional Review Board and Research Compliance Services (Project ID: 18410, IRB Approval #: 20180618410 EX). Coefficients for internal consistency for the primary estimation variables varied (Appendix B), and weaknesses were actively addressed.

The first pilot instrument was compiled using information from similar instruments and influenced by literature about self-efficacy, the Opportunity to Learn concept, and how Educators' Perceptions of Their School Environment and school characteristics potentially impact achievement. With considerations for survey efficiency (administration, parsimony, reliability, and cost), the instrument development process was also attuned to issues such as survey length (number of items), structural formatting, time investment, and content or construct coverage.

The instrument was developed for paper-based administration because of limitations associated with telecommunication services and access to computers in the district. In Belize, there were relatively high costs associated with the printing and distribution of surveys. Surveys of this nature were not commonly used in the education system; therefore, a lack of familiarity with the process or the value of participation was also anticipated and addressed in the survey design. The study sample was also limited to primary school educators. The instrument was culturally adapted for Belize. Feedback from multiple sources informed item revisions. Care was also taken to ensure that the language used was culturally sensitive and appropriate for all educators (interns, teachers, and administrators). Decisions about the design format included limiting the instrument to three double-sided pages and using visual cues and white space to minimize effort and time commitment. It was also crucial that the instrument capture as much information as possible in a single application. A concerted effort was made to encourage participation by showing professional relevance and explicitly presenting the survey as an opportunity to include the educators' perspective or voice. The processes involved local partners and the Belize Ministry of Education. Visible measures were taken to secure and respect confidentiality. All educators had the option to enter the sweepstakes drawings for 2 Kindle Fire-7 tablets and a unisex handmade Arabian Satchel as part of the researcher's recruitment efforts.

The documents associated with the IRB approval of this study are included in Appendix C (Appendix C1-C6). After the necessary analyses and revisions, the government of Belize granted formal permission to conduct the *Belize Educator Survey*. Thereafter, the University of Nebraska Institutional Review Board and Research Compliance Services (IRB) approved data collection in Belize using the *Belize Educator Survey* (Project ID: 18717, IRB Approval #: 20190118717 EX). The primary sources of data for this study came from the revised version of the *Belize Educator Survey* (Appendix C4) and data provided by the Belize Ministry of Education and District Education Office. Appendix D presents general school statistics for the district in which the *Belize Educator Survey* was distributed. This *Belize Educator Survey* was deemed culturally appropriate and intended only for educator samples in Belize.

Items

Bandura's (2006) suggested guide for constructing self-efficacy scales supports the use of Likert-type response formats using ranges in which participants record the strength of their efficacy beliefs on a scale from 0 to 100 with 10-unit intervals. Bandura also noted that "a simpler response format retains the same scale structure and descriptors but uses single unit intervals ranging from 0 to 10" (p. 312). In this study, a 10-point Likert-type scale was used, and the response options ranged from 1 to 10 (Appendix C4). Items in each section were organized so that participants would consider each item independently before selecting or supplying a response. To optimize clarity, response options were deliberately parsimonious and positively coded. Supplied responses requested numbers only.

Relevant existing sources with scales pertaining to Educator Efficacy and Educators' Perceptions of Their School Environment were adapted for the present study. Preexisting sources were the Educator Efficacy Scale (Bandura, 2014) and the *Lincoln Public Schools (LPS) Teacher Perception Survey*. Bandura's Educator Efficacy Scale was previously adapted for use in Belize in an IRB-approved study for educators preparing students for the Primary School Examinations in Belize (Usher-Tate, 2014, 2015). That version was further revised and incorporated in the present instrument. The LPS in Nebraska administers its perception-surveys to students, educators, and parents. The LPS measure for Educators' Perception of Their School Environment in Lincoln was used internally; the results were not published.

In the absence of a relevant existing and freely available indicator for the Opportunity to Learn construct, the items for the Opportunity to Learn construct were developed, not adapted. The foci for items addressed the selected aspects of the Opportunity to Learn construct and were generally related to access, context, and resources that were addressed in cross-national studies and educational research reviewed for this study (Agasisti et al., 2018; Chaudry & Wimer, 2016; Chin & Chow, 2015; Heafner & Fitchett, 2015; Hurtado, 2001; Näslund-Hadley et al., 2013; O'Malley et al., 2014; Schuh-Moore et al., 2012; Segretin et al., 2016; Stephens et al., 2015).

Cultural Appropriateness Review

The *Belize Educator Survey* instrument was reviewed before the pilot administration (June 2018). Two graduate students reviewed the instrument for survey structure, timing, and clarity. Four Belizean educators and three Belizean college-level students (all of whom were originally from Belize and had attended public schools within that region) reviewed the instrument for sociocultural linguistic nuances, clarity, and relevance. The Belizeans also provided feedback on the language used and the time allotted to complete the survey. A psychometrics professor also read the materials and offered suggestions for structural adjustments. The pilot materials also included a survey feedback sheet, which most participants did not use. Post-pilot administration, two graduate students of Latino heritage, two Belizean college-level students, and seven Belizean educators reviewed the instrument and provided additional feedback. Adjustments were made to the survey instrument in consideration of feedback, item statistics, and response patterns.

Given the limitations associated with technology in the parts of rural Belize and the need for consistency, it was most appropriate to continue with the paper-and-pencil administration only. The pilot administration sampled one primary school with 22 educators, of which 21 returned responses (95.45% response rate). The internal consistency coefficients for estimation variables in the pilot study were used to inform the revision process (Appendix B) for the instrument used in the present study.

The Educators' Voice

This study was intentional in seeking the voice of educators through the two educator-related variables: Educator Efficacy and Educators' Perceptions of Their School Environment. The items that composed the variable for the Educators' Perceptions of Their School Environment sought educators' standpoints on multiple issues related to the learning environment, educator influence, educator context, school climate and culture, curriculum and assessment, and student engagement. Item analyses on the pilot data confirmed the anticipated relationships for the two educator-related indicators. Given the feedback from the pilot, the number of items in both indicators increased by two to reduce compound statements to a single focus in each. For example, one of the items in the pilot instrument read, "How much are you able to keep students on-task with difficult tasks or to stay focused during difficult times?" The revised form separated the two issues. "How much are you able to keep your students 'on-task' with difficult tasks?" and "How much are you able to keep your students 'focused' during difficult times?" One item was also introduced to bolster the content coverage of curricula. "How much are you able to make sure that your students learn grade-level content?" Other revisions to the Educator Efficacy section were mostly cosmetic.

The Educator Efficacy variable also had Likert-type items with options ranging from 1 to 10. For this estimation variable, 1 was the lowest level (negative impression), and 10 was the highest or strongest level (positive impression). The internal consistency reliability (coefficient alpha) of the pilot version was very satisfactory (α = .909) for the Educator Efficacy variable. Principal components analysis with oblique rotation extracted three components that accounted for 73.51% of the total variance. The selected themes for Educator Efficacy components were control of the learning environment, personal influence, and professionalism.

The pilot version of the Educators' Perceptions of Their School Environment used the five response options format (strongly disagree, disagree, agree, strongly agree, and "I don't know"). No participant used the "I don't know" option included; as a result, that option was not included in the revised version. The coefficient alpha was satisfactory (α = .800). A principal component analysis perception estimation of the variable with an oblique rotation extracted nine components that explained 86.44% of the variance in the variable. The items in the Educators' Perceptions of Their School Environment were intended to inform three subparts or themes relevant to the school culture and climate, curriculum and assessment, and engagement. Item revisions for this estimation variable were minor and included revising the language (word choice and clarification) in four items and splitting the content of two items. For example, one statement in the pilot instrument was "Teachers and administrators in this school emphasize effort." In the revised instrument, that statement was broken into two separate statements: "Teachers in this school emphasize effort," and "Administrators in this school emphasize effort." The final version of this estimation variable had 29 items, and to be consistent, the response options were presented on the 1-to-10-point Likert-type scale, in which 1 was *strongly disagree* (negative impression) and 10 was *strongly agree* (positive impression).

Opportunity to Learn and School Context

Additional feedback was sought to improve this part of the instrument because the internal consistency coefficients in the pilot were not acceptable (α = .574) for Opportunity to Learn and (α = .605) School Context. Suggestions and comments were directed to the language, clarity of intent, and formatting. The items that were flagged for clarification were revised for clarity or to increase specificity. A few educators also suggested including items that specifically addressed the prevalence of poverty (in terms of children coming to school hungry) and students who were only exposed to English in school when the adults (parents/guardians) at home did not speak English.

Some of the targeted issues for the Opportunity to Learn and School Context estimation variables included, but were not limited to, access, resources, and educator qualifications, years of service, school size, management type, location, and physical condition. The Opportunity to Learn and School Context portions of the pilot instrument failed to assess the constructs reliably. The internal consistency reliability coefficients were much lower than anticipated. Fundamentally, the issue was that items were not as narrowly focused (Wainer & Thissen, 1996). For the most part, revisions of the survey instrument were concentrated on the Opportunity to Learn and School Context segments, resulting in new items, revised items, and adjustments for structural coherence. With the feedback and considerations, the *Belize Educational Survey* was changed to include 23 more narrowly focused items for Opportunity to Learn and 10 for School Context.

Academics

In the *Belize Educator Survey*, educators' reports of their classes' averages were used to formulate the School-Based Averages variable. School-Rigor was quantified as the differences in overall school-level performance on internal and external assessments. School-based academic performance was computed using statistics reported in the survey for the core subjects (English, Mathematics, Science, and Social Studies). The School-Rigor variable was computed using observed scores for the 2018 Primary School Examinations, courtesy of the Belize Ministry of Education, and the School-Based Averages. The two components of School-Rigor were described in terms of internal rigor and external rigor. Internal rigor was the difference in school-mean for School-Based Averages and official school-mean for the 2018 administration of the Primary School Examinations. External rigor was the difference between official school-mean and the district-mean for Primary School Examinations.

Conceptual Model

Conceptually, School-Level Achievement is the product of contributing factors and variables associated with the three main bodies involved: the school, the educators, and the students (Figure 1, p.5). Therefore, School-Level Achievement encapsulated variables that accounted for the educators' voice (Educator Efficacy and Educators' Perceptions of Their School Environment), school opportunities and characteristics (Opportunity to Learn and School Context), and academic performance (School-Based Averages, School-Rigor, and Primary School Examinations). Educators' voices were measured in terms of responses about Educator Efficacy, Opportunity to Learn (access and resources), and Educators' Perceptions of Their School Environment. School Context and School-Based Averages were educator-level or class-level estimation variables. The educator-level estimation variables were computed using data from the *Belize Educator*

Survey. The school-level estimation variables included the other academic performance estimation variables, School-Rigor and Primary School Examination–related scores.

From a traditional education standpoint, the estimation variables in this study were analogous to subject or course averages. Much like subject averages, scores for each variable provided information about that area of focus, a valued component of the collective description of achievement. In addition to the information collected in the *Belize Educator Survey*, the Belize Ministry of Education provided data about school-level characteristics.

Sample

There are six districts in Belize, and each district has a district-level education center that guides and supports all primary and secondary schools (K–12) (Government of Belize, 2003; Belize Ministry of Education, 2012a; Belize Ministry of Education, 2012b). The present research study collected data from schools supervised by the same district education office. This school district was culturally diverse. Census statistics showed noticeable shifts in the demographics of this district in recent decades. Approximately half of the primary school students in this district attend schools managed by the Catholic diocese. Other churches (Adventist, Anglican, Assembly of God, Evangelical, Mennonite, Methodist, and Nazarene), a particular community, or the government manage the remaining schools (Usher-Tate, 2015).

Full-Coverage Sampling

Lower response rates are typical for voluntary surveys; therefore, full-coverage sampling methodology was most appropriate (Knaub, 2011). The total population of this district was relatively small. In the present study, all primary school educators in the district were invited to participate.

In the 2017–2018 academic year, 39 primary schools operated within the jurisdiction of that particular district: 38 government or church-state public schools and one private school. The 38 public schools were located in three districts but were managed by the same district education office. These schools were in District A (n = 32), District B (n = 1) and District C (n = 6). Other demographic information sought included: availability of preschool component, school size (number of students and educators), and management type (church, government, community). Location categorization as an urban, rural, or remote area involved considerations for distance from the main district centers, infrastructure, and access to public transportation. The instrument, strategies, and protocols used in the present study appear appropriate for replication in other districts in Belize and nationally.

Power

The "a-priori" power analyses using G-Power 3.1 indicated that to achieve a strong effect size (0.05) with three groups, 0.95 power (1-β) and 0.01 error (α) probability. The probability of a correct decision with those parameters required a sample size of 252 participants. The goal in this study was to have 250 educators participate to let the study go forward.

Data Collection Process

The data collection portion of this project was funded by the Lois E. Fellowship Award, a prize donation from a professor, and the author's personal funds. Contributions from the Government of Belize included sharing data and coordinating the survey implementation. As per IRB-approved protocols, the secured internal distribution and retrieval of survey were managed through the established network at the district education office.

To create awareness in the school district, flyers (Appendix C2) about the project were distributed to all primary schools in the district. As per IRB-approved protocol, in the two weeks prior to distributing the survey, personnel from the Belize Ministry of Education (District Education Office) also contacted school principals via email. Contact information for schools was provided to the principal investigator to share the flyers electronically and to allow for timely responses to direct questions.

Participants sought in the data collection phase of this study were educators (teachers and administrators) who worked at a primary school (grades one through eight) in the school district during the 2017–2018 academic year. In the Belizean context of education, especially in the school district of interest, most primary schools do not have secretaries, and many principals have dual roles as an administrator and a full-time teacher. In these schools, there were no mailbox systems in which to put the envelopes. School packets included a flyer that summarized the intent of study; a large, addressed, polyurethane return envelope; and individually packaged survey materials for all educators at the school. Each individual survey packet also contained an informed consent notice (Appendix C3). This notice served as a further reminder that participation was voluntary and that there were no consequences for choosing not to participate. In addition, individual packets also included a survey instrument, a sweepstakes ticket, an envelope labeled with instructions (Appendix C4), and the survey instrument (Appendix C5). All educators at the school were invited to participate and all were equally eligible for the sweepstakes drawings for two Kindle Fire-7 tablets and a unisex handmade Arabian Satchel. More than a dozen educators who did not complete the survey submitted a completed sweepstakes ticket that were included in the draw. There were also educators who completed the survey and opted not to participate in the sweepstakes.

The individual survey packets were distributed to all educators (teachers and administrators) via their established network. Involving the partners in the established network circumvented any appearance of “targeting” or “favoritism.” An established network also helped to ascertain accountability for survey instruments. In the accompanying instructions, educators were asked and reminded to seal their envelopes and return them to the principal or the person designated to collect the envelopes at their school. Sealed envelopes assured school- and district-level confidentiality and protected the individual’s participation status from others. Once recovered, the sealed envelopes were consolidated into a single package and shipped via registered express mail to the lead researcher in Nebraska. Sufficient completed surveys were returned such that the study could proceed without additional follow-up measures. The day after the author received the packets, winners were selected and contacted. The three winners collected their prizes within two days of notification.

Base-Model

Items to Estimation Variables

As in most educational settings, the general connotation was that higher scores are better and lower scores are not as good. Therefore, all item scores derived from items with non-positive connotations were reverse coded. The reverse coding of such items maintains uniformity with scoring so that higher values for each item consistently represented positive connotations and lower values were less favorable. Item scores were then grouped according to content mapping or alignment with the concepts that were subsequently referenced as estimation variables.

Section one of the *Belize Education Survey* had questions on a variety of issues. These items mapped separately onto the concepts of Opportunity to Learn and School Context. The second section of the survey contained items about the Educators' Perceptions of Their School Environment. In section three, educators rated their efficacy levels on work-related items. The final section of the survey had items that mapped on to Opportunity to Learn, School Context, and School-Based Averages.

In addition to data from the *Belize Educator Survey*, the Belize Ministry of Education also provided contextual information about school sizes (number of students and number of educators), locations, and school managements. In total, there were 10 items mapped onto School Context, 20 items onto Educator Efficacy, 23 onto Opportunity to Learn, 29 onto Educators' Perceptions of Their School Environment, and 4 onto School-Based Averages. The cumulative scores for each concept were essentially the preliminary versions of the estimation variables used in this study: School Context, Educator Efficacy, Opportunity to Learn, Educators' Perceptions of Their School Environment, and School-Based Averages.

Uniform Score Range

There were differences in some item structures and the number of items across sections for the various estimation variables included in the survey. Therefore, total combined scores (sums) for each estimation variable were converted to have a score range of 0–500 points for comparability (Suhr & Greeley, 2004). The transformation or conversion of the score values were based on possible outcomes per variable and not confined to the range of observed scores. In the transformation process (formula), the educators' response was *E*, the lowest possible score for the estimation variable was *L*, the highest possible score was *H*, and the new score value use across all estimation variables in this study was *N (500)*. The formula to transform or rescale was (*E*-*L*) / (*H*–*L*) * *N* (Suhr & Greeley, 2004). The typical score range used in educational settings is 0 to 100; this study used 0 to 500 as one means of highlighting differences while maintaining some level of systemic familiarity with educational grade computation and records.

Missing Data

For the most part, across the 36 schools represented in the data, the 294 educators who participated completed all sections of the survey with very few missing responses, except for the section that requested class averages, for which there were 46.9% of teachers who did not report grades. Multiple educators made notes on their instrument that they were unable to provide this information because they no longer had access to their grade books from the previous year. First, the nature of this study was to examine two methods of estimating school-level achievement. Second, the expectations were such that if these methodologies were employed by school districts, the survey would be mandated, not voluntary, and the educators would be providing current grades, with fewer instances of missing grades. Instead of deleting this variable, imputation or simulation of the missing data was imperative, and caution thus noted for inferences based on the variable for School-Based Averages.

Statistically, simple mean imputation for missing data preserves the mean structure of observed data at the school level and enables retention of sample size. In other words, if missing scores were replaced with the school mean for an estimation variable, the school's overall mean for that particular variable would not change, and all the educators who had a missing score were retained in the sample. One benefit was that information provided by those educators was also retained (each educators' voice was still included in the school-level estimate). However, a major disadvantage of using simple mean imputation was the weakening of within-variable variability, a loss in the richness of information, or overgeneralization. Therefore, simple mean imputation would not be advisable if inferences or decisions were anticipated at the educator level.

To retain school-level observed means (the unit of reporting in this study), the preferred alternative to the simple mean imputation was stochastic (random) imputation with weighted probability. With stochastic imputation, random numbers were generated, which means more variability in the data than what was likely for imputations using the school means (Scholtus et al., 2014). The random numbers were generated using a set of parameters (weighted probability), and the combination retained the mean structure because the numbers generated were not entirely arbitrary (Scholtus et al., 2014; Seaman et al., 2012). The numbers generated were within a probable range determined by the observed response pattern. Stochastic imputation with weighted probability retained the mean structure with more variability than simple mean imputation.

Unlike simple mean imputation, stochastic imputation with random weighted probability is a multistep process. In this approach, a helper table was created for each of the four subject areas (English, Mathematics, Science, and Social Studies). The helper table listed a range of observed scores by intervals, the observed count per interval, the probability (likelihood) of observing a score in each interval, and the cumulative percent. The helper tables shaped the parameters for the weights (probability and cumulative percent). Fundamentally, the randomness of the numbers generated (RAND) was weighted by the parameters of the helper table, linked to an interval score (MATCH), and limited to a specified range (Burns, 2019). To compose each number that was eventually used to replace a missing number (imputation) for a given variable, the process used in this study generated the components of each number a single digit at a time. The first digit generated was designated as the initial digit (e.g., the 9 in 92.35, or 6 in 67.51). This digit was random within the constraints imposed by the probability factor. The second digit (e.g., the 2 in 92.35, or 7 in 67.51) and decimal values (e.g., the .35 or .51) were generated using syntax that did not include probability or the helper table. Each imputation value was compiled using the digits in the order generated. Microsoft Excel® has the capacity to generate these values using the RAND and MATCH functions. This is an example of the syntax (MATCH(RAND(),Q$3:Q$8)) for the first digit where Q$3:Q$8 referenced the helper table. For the second digit, there was a random number between zero and nine (RANDBETWEEN(0,9)). The syntax for the decimal value limited the random generation to numbers greater than zero but less than one.

The stochastic imputation with weighted probability process was repeated (all steps) for missing values in each subject area. A comparison of the district-wide and school-level descriptive statistics for the observed data and the data set with imputed values showed very similar means and medians for each subject area.

School Types

To determine if there were significant relationships between school management and achievement, schools were grouped according to managing authorities. The responsibility for the operational management of public primary schools in Belize falls into one of three categories of school managers: Catholic, other denominations, and Non-parochial (government and community). The presence or absence of an incorporated or affiliated preschool was also considered as another option for classifying public primary schools.

Descriptive Statistics

To minimize the introduction of confounding variables, the Base-Model was used as the foundation data set for both estimation models tested in this study. The Base-Model data set was described in terms of the mean, mode, median, range, frequency counts, standard deviations, and standard error of the mean. The Base-Model data set included observed district-level and school-level data sets. The district-level data set referred to the responses of all educators, not constrained by school affiliation. School-level data sets were compiled using educator responses clustered by affiliated schools. The profile and quality of the Base-Model data set were confirmed after the data cleaning activity. General descriptive statistics and reliability coefficients were estimated for each of the estimation variables: School Context, Educator Efficacy, Opportunity to Learn, Educators' Perceptions of Their School Environment, and School-Based Averages.

Models: WISP and MAEP

The two methodologies for estimating school-level educational achievements were Weighted-Indicator Scores Protocol (WISP) and a Multilevel Achievement

Estimation Protocol (MAEP). Both methodologies were considered because of their appropriateness for applications that involved data with the common issue of small and unbalanced samples. Unbalanced sample sizes are commonly associated with data from organizations and schools as unit structures. Given the variety of school sizes and types in this school district, unbalanced sample sizes between schools were anticipated. School-level-only estimation and reporting intended for low-stakes decisions minimized potential disadvantages that may otherwise have been associated with limited variability, especially for schools with small representation. To such schools (smaller/low representation), these two methodologies allowed the opportunity (benefit) for inclusion in the overall estimation and the option to give schools meaningful feedback. Maximum likelihood (ML) estimators in MAEP also made it possible to estimate unbalanced groups in multilevel-multivariate conditions (Hox et al., 2010).

Weighted-Indicator Scores Protocol (WISP)

The WISP method encapsulates a protocol (set of procedures and analyses ordered for this study) with measures and analyses appropriate for small samples in educational settings. Base-Model scores were transformed using the Weighted-Indicator Scores Protocol, and the result was an overall estimate of school-level achievement. Comparatively, in schools, educators submit subject-area grades used to compute grade-point averages (GPA). GPAs generally represent overall student achievement. Similarly, the WISP collects information from educators to establish the scores for each estimation variable. Like subject-area grades, the scores for each estimation variable independently describe the foci. The estimate of the collective value of the estimation variables describe overall school-level achievement, like term grades or GPA do for students.

Data Quality

Base-Model estimation variables underwent a series of analyses to establish how suitable the variables were for the Weighted-Indicator Scores Protocol. Assessment of the internal consistency was the first analysis to evaluate the reliability of the estimation variables. The first criterion was that the overall scores for the estimation variables had to be comprised of four or more items or data points. The internal consistency of each estimation variable was estimated independently to determine corresponding coefficient alphas (α).

An analysis of variance test (ANOVA) assessed whether there were significant differences in the variances across the estimation variables. With ANOVA procedures, assumptions for normality of distribution and homogeneity of variance were first satisfied. The Shapiro-Wilk test was used to assess normality in the distribution. The Levene statistic tested for homogeneity of variance. The correlational relationships between the estimation variables were also examined and expressed in terms of Pearson correlations and significance tests.

Once the quality of the data distribution was satisfied, the next step used the educator-level Base-Model data to give all educators a voice in setting the weights (n = 294). In that process, the data were further assessed using the Kaiser-Meyer-Olkin Test (KMO) and Bartlett's Test of Sphericity.

KMO

KMO tests measured sampling adequacy and determined the proportion of variance in variables attributed to underlying factors (IBM Corporation, 2019). For the

KMO test, values closer to 1.0 are desirable because values less than .50 indicate that the results obtained in any subsequent factor analysis most likely will not be meaningful.

Bartlett's

The Bartlett's Test of Sphericity assessed correlations among items comprising the estimation variables (CTX, EFF, OTL, PRC, and SBA). If items for a variable are uncorrelated, then they do not have a structure to assess and should not be combined to create an estimation variable for the study. With the Bartlett's Test of Sphericity, values less than .05 indicate that (with that level of significance), it is not due to chance that a factor analysis conducted on the data could produce meaningful results (IBM Corporation, 2019).

Confirmatory Factor Analyses and Weights

Confirmatory factor analyses (with varimax rotations and maximum likelihood extraction) were conducted separately for each estimation variable. The decision regarding the number of factors retained for each variable also took into consideration two traditional approaches, Kaiser's rule and Cattell's scree-test. Kaiser's rule, or the eigenvalues-greater-than-one rule (K1), is a default extraction or decision rule in the SPSS software, and retained for this study. Cattell's scree-test takes a more visual approach to examine the scree plot for logical and visual cutoffs (Cattell & Voglemann, 1977). That is, viable factors are those that precede where the slope tapered off or flattened.

Similar to other seminal or classic theories, published before the widespread use of computing technology, together the Kaiser's rule and Cattell's scree-test were the commonly used metrics to make sense of data (Ledesma & Valero-Mora, 2007; Nelson, 2016; O'Connor, 2000; Reckase, 1979; Zopluoglu & Davenport, 2017). Studies have shown that using Kaiser's rule and Cattell's scree-test tend to overestimate number of factors (Ledesma & Valero-Mora, 2007; O'Connor, 2000). However, factor reduction in the traditional sense was not the purpose of assessing dimensionality or the number of factors in each variable in the Weighted-Indicator Scores Protocol (WISP). The results from Kaiser's rule and Cattell's scree-test were used to corroborate the alignment to the content foci, not to reduce, or differentiate, or identify the specific contributing factors within each variable. Estimation variables were used as composite scores in this study.

The WISP considered factors extracted as viable in the determination of the weight value only if the eigenvalues corresponding to the first factor accounted for more than 20% of the total variance (Reckase, 1979; Zopluoglu & Davenport, 2017). The total percentage of variance accounted for by factors with Kaiser's rule were incorporated into to determination of weights. The rationale was that the total percentage of variance explained how much of the variance was associated with the construct of the estimation variable. WISP weighted values were computed separately using the total variance explained and the first eigenvalue. The weights were applied to Base-Model scores for variables, resulting in the scores for the WISP estimation variables at the district and school levels. The weighted scores used in the WISP estimation were considered analogous to subject-level grades.

The WISP analyses attended to the data quality and further extended the Base-Model. Weights were determined by the variances explained in the factor analyses. The District-level and School-Level Achievement estimates were computed using values from the weighted scores and the value for School-Rigor, much in the same was as educators applying weights to various components in a course to determine a final grade. Subsequently, the estimates produced using WISP were evaluated and described in terms of the mean, mode, range, standard deviations, and the standard error of the mean.

The Multilevel Achievement Estimation Protocol (MAEP)

Overview

A multilevel structural equation model (MSEM) approach was used to estimate multilevel multivariate data (Hox et al., 2010). The MSEM approach took into account the issues and benefits possible with both structural equation modeling and multilevel modeling (Gelman, 2006; Hox et al., 2010; Rabe-Hesketh et al., 2007). The MSEM framework also provided unbiased estimations of level-one variables on other level-one variables (Preacher et al., 2010).

Estimation models using MSEM were based on data from independent variables that were collected at various nested or cluster dependent (nonrandom) levels (Muthén, 1991; Rabe-Hesketh et al., 2007; Tabachnick & Fidell, 2013). Multilevel models take into account that data were nested and that sampling occurs at two levels (Muthén & Muthén, 1998–2017). This kind of structural equation modeling is appropriate for studies involving hierarchical organizations such as education systems (Hox et al., 2010; Kaplan & Elliott, 1997), where the relationships can also be complex (Kaplan & Elliott, 1997), as was the case for this study.

The MAEP Model

The Multilevel Achievement Estimation Protocol (MAEP) used in this study was an adaptation of the “two-level SEM with continuous factor indicators and a random slope for a factor” example in Muthén and Muthén (1998–2017, pp. 228–231). The school-level educational achievement was estimated with the methodology applied within the context of parameters specific to this study and depicted in Figure 2. Level-one data were the educator-level data, which were nested in level-two data at the school level. In the level-one (educator-level) part of the model, the estimation variables were continuous scores boxed because they represent observed scores. The filled circles at the end of the arrows from the latent construct Achievement Within School to the observed estimation variables represented random intercepts. The intercepts were subsequently presented as latent variables in the level-two (between-clusters) part of the model.

In the MAEP framework graphic (Figure 2), the random intercepts for School Context, Educator Efficacy, Opportunity to Learn, Educators' Perceptions of Their School Environment, and School-Based Averages were encircled because they represented continuous latent variables that varied across the schools. Subsequently, School-Based Averages and the random intercept Educator Effect/Bias were regressed on the latent construct for achievement and School-Rigor in the level-two part of the model.

In Figure 2, the within-school-level (level one) boxes represented unweighted, observed, continuous, dependent variables. Latent variables included AWS in level one and Achievement Between Schools in level two. School-Rigor was the only observed continuous independent variable in level two. Much like the intercepts, the random slope was introduced (small filled circles) in level one and computed in level two.

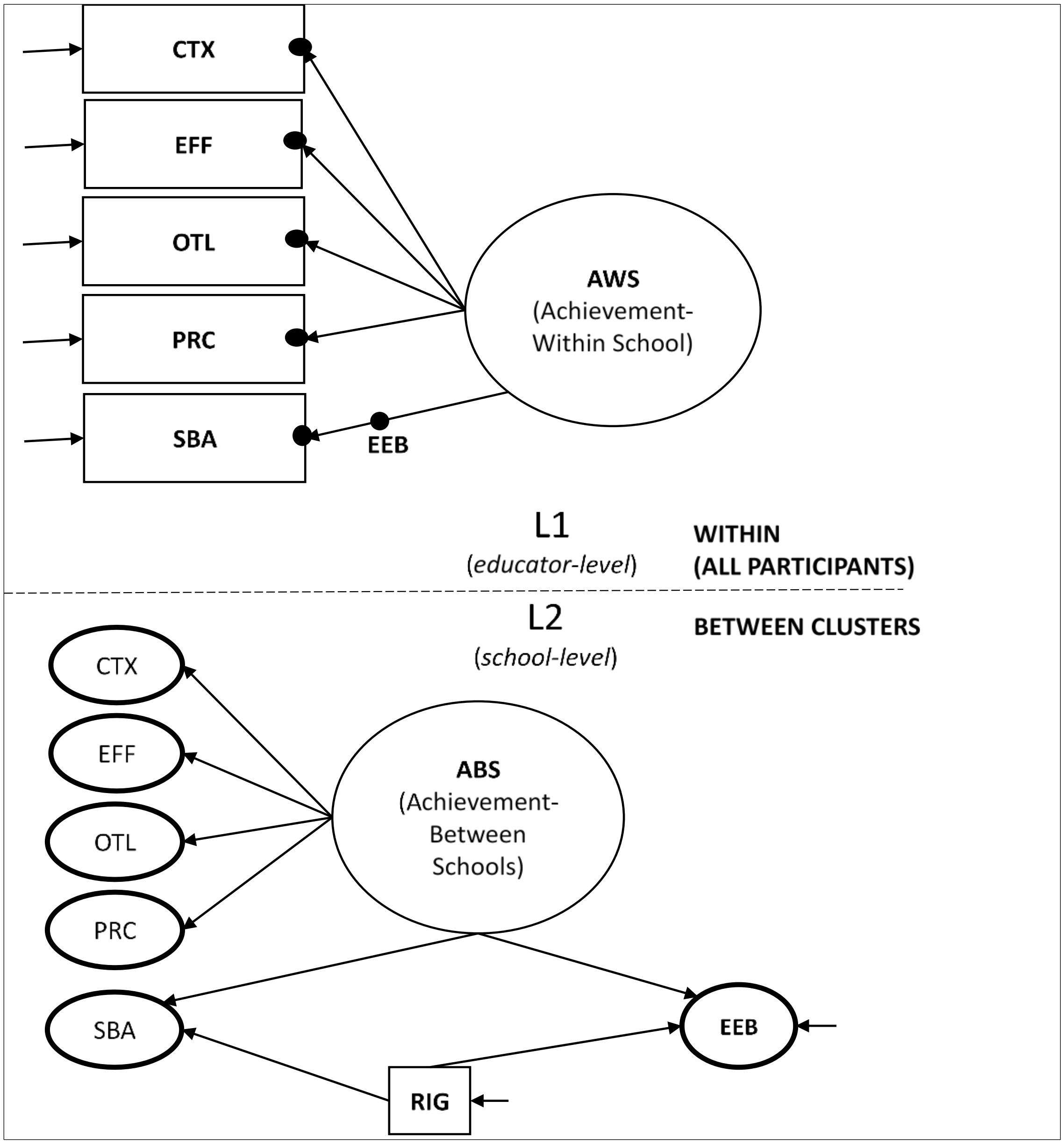


Figure 2. Multilevel Achievement Estimation Protocol. In this two-level structure equation model used in the Multilevel Achievement Estimation Protocol (MAEP), the observed continuous estimation variables were boxed in level one: School Context (CTX), Educator Efficacy (EFF), Opportunity to Learn (OTL), Educators' Perceptions of Their School Environment (PRC), School-Based Averages (SBA), and School-Rigor (RIG). The latent factors were presented in the larger open circles: Achievement Within School (AWS) and Achievement Between School (ABS). Level-one random intercepts (solid, filled circles) that vary across schools were represented as open circles in level two. The random intercept was Educator Effect/Bias (EEB).

Estimation

In the MAEP in this study, the data set for the MSEM portion of the analyses was first converted into a text file format so it could be read into the M*plus*® software. For MSEM analyses using M*plus*®, the data set was defined, and all variables were specified in the input syntax for this particular software including the cluster and latent variables (see summary of syntax in Appendix E). The second part of the M*plus*® syntax described the analyses specified in the model (Figure 2). The third and final aspect was the last line of syntax to indicate the type of output desired. The analyses for correctly specified models terminated without warning, and the output became available in a single step. The computational speed of the M*plus*® software and the simplified specifications for the model in this study processed very quickly (less than two minutes).

The model results for the multilevel structural equation model estimation using M*plus*® reported four statistics: the parameter estimate, the standard error (SE), the parameter estimate divided by the standard error (Est./SE) and *p*-values. "This statistical test (Est./SE) is an approximately normally distributed quantity (*z*-score) in large samples. The critical value for a two-tailed test at the .05 level is an absolute value greater than 1.96" (Muthén & Muthén, 1998–2017, p. 798). The overall value used for the MAEP approach was the sum of the parameter estimates divided by the standard error for achievement between.

MAEP Descriptive Statistics

The descriptive statistics (school and district levels) were computed for the estimates derived from the MAEP approach in this study. The estimates were evaluated and described in terms of the mean, mode, range, standard deviations, and standard error of the mean.

Research Questions (Analyses)

The first question in this study was: *Is there a difference in the overall School-Level Achievement estimates using a classic summative approach with the Weighted-Indicator Scores Protocol (WISP) versus a latent approach with a Multilevel Achievement Estimation Protocol (MAEP)?*

The first step in answering this question was to standardize and compare the descriptive statistics for the Primary School Examination, Base-Model, WISP, and MAEP estimates. Then overall estimates were sorted in descending order and ranked. Model estimates were matched with rank-ordered Primary School Examinations estimates. A match with the Primary School Examinations rank-ordered estimate counted when the school rank was exactly the same as the Primary School Examinations rank or off by one or two ranks. For the overview of the model estimates, the overall statistics for the district and a few school examples were compiled by estimation variable and model. To determine if there were differences between the school-level estimates for WISP and MAEP, a *t*-test (.05 alpha) was conducted. Pearson correlations estimates were also examined, and a multiple linear regression analysis was used to predict Primary School Examinations on the basis of the Base-Model, WISP, and MAEP school estimates.

The second question in this study was: *Were there differences in the School-Level Achievement estimates by selected school characteristics such as School Type (Catholic, Non-parochial, or Other Denominations) or Location (Urban, Rural, or Remote)?* The analyses conducted were two one-way ANOVA with Tukey HSD post hoc tests. These analyses tested for differences in model estimates across schools when a categorical variable was considered (Location or School Type). Computed eta squares (η^2) indicated the amount of variance associated with the main effects.

Question three in this study was: *How much variance in school-level academic outcomes was accounted for by selected Educator-Efficacy, Educators' Perceptions of Their School Environment, and the Opportunity to Learn variables?* For the third question, three independent multiple linear regression analyses were used to test how well the Educator Efficacy, Educators' Perceptions of Their School Environment, and Opportunity to Learn variables predicted the academic outcomes: School-Based Averages, Primary School Examinations, and School-Rigor. In the first regression, the dependent variable was Primary School Examinations; in the second, the dependent variable was School-Based Averages; and in the third, School-Rigor. The predictor variables were Educator Efficacy, Educators' Perceptions of Their School Environment, and Opportunity to Learn in each of the three analyses.

Question four in this study was: *Was there a difference between class averages for upper, middle, and lower divisions and school-level performance on the 2018 Primary School Examinations?* For the fourth question, a one-way ANOVA with Tukey HSD post hoc (where necessary) was used to compare differences between class-level divisions (upper, middle, lower) for School-Based Averages and Primary School Examinations. Lower division was defined as Infants 1 and 2 (grades one and two). Middle division included Standards 1, 2, and 3 (grades three through five), and upper division included Standards 4, 5, and 6 (grades six through eight).

Question five in this study was: *What was the relationship between selected school characteristic variables (School Size, Location, and Infrastructure) and variables for Educator Efficacy and Educators' Perceptions of Their School Environment?* For the fifth research question, Pearson correlations (*r*) were used to compare relationships between educator-related estimation variables and school-characteristics.

Chapter 4

Results

The results of this study are presented in three parts. The first section (Data) describes the findings of the process and data associated with the Belize Education Survey. The second section (Estimation) reports the statistics for the Base-Model and the two subsequent approach models: Weighted-Indicator Scores Protocol (WISP) and Multilevel Achievement Estimation Protocol (MAEP). The third section (Questions) of this chapter addresses the results of analyses used to answer the five research questions of this study.

Data

There were 524 survey packets provided for distribution. After reconciling the returned survey packets, the return rate was approximately 60.11% (n = 315 sealed responses). Envelopes that were not sealed and had no marks or sweepstakes entries (n = 21) were not included in the count. Therefore, the final count included 294 (56.1%) returned packets, representing 36 schools. Additionally, 19 educators declined with an active response or no response but submitted a completed sweepstakes ticket. Overall, there were 44 more active participants than the minimum (250) determined a priori for the study to proceed.

Educator and School Count

Of the 294 participants, 28 (9.52%) educators had transferred to a different school for the 2019–2020 academic year. The educators with transferred status indicated that they completed the survey based on the school relevant to the 2017–2018 academic year. There were also three schools represented entirely by educators who had transferred out of those schools. The final school-level sample count included representation from 36 of the 38 (94.7%) public schools in the district. One recently established school, included in the count, only had preschool through third-grade classes in the 2017–2018 academic year.

Location

The physical distance between schools and the District Education Office ranged from 0.15 kilometer (0.09 mile) to 97.6 kilometers (60.65 miles) with average estimated driving times ranging from 2 to 95 minutes (considering road types and access in normal conditions). Location categories (Urban, Rural, and Remote) were designated in consideration of status attributed by the Belize Ministry of Education, distance from one of two main district centers, community infrastructure, and access to public transportation. In the data set, the representation included the following school counts by Location: Urban (n = 92), Rural (n = 115), and Remote (n = 87).

Size and Type

The Belize Ministry of Education provided school-level statistics about structure and composition (Table 1). There were 38 public primary schools in this school district, with 885 preschool students and 9,128 primary school students in grades one through eight. The average number of primary school students per school was 240, with approximately 30 students in each class.

School Type depended on the entity responsible for the day-to-day operation of the school, either church-state or Non-parochial. Further differentiation in the church-state schools was binary—Catholic or Other Denominations (Table 1). The Other Denominations School Type included schools managed by Anglican, Methodist, Adventist, and Evangelical leadership. The Non-parochial School Type included schools managed by the government or a community.

Table 1

General Demographics by School Type

School Type (description)	*n* Schools	*n* Students	*n* Educators	*M* Student-Teacher Ratio	*n* Preschools
Catholic	11	3800	184	21.68	10
Non-parochial*	11	2894	129	22.09	10
Other Denominations**	14	2659	135	21.83	10
Overall (district)	36	9353	448	21.86	30

* Includes schools managed either by the Belize Ministry of Education or community-based organizations (not church-affiliated)

** Includes schools managed by leadership in the following churches: Seventh Day Adventist, Anglican, Assembly of God, Baptist, Church of Christ, and Methodist

The Other Denominations School Type had the most number of schools ($n = 14$) but had the least number of students ($n = 2{,}659$) and educators ($n = 135$). The Catholic and Non-parochial entities managed an equal number of schools ($n = 11$). However, the Catholic management was responsible for the largest number of students and educators in the district.

Access and Allocation

Three school-characteristic variables discussed in earlier chapters lacked variability across schools: the student-teacher ratio, availability of preschools, and time allocated for core subjects.

The number of students to teachers (including administrators who also teach) was the student-teacher ratio (Table 1). The observed differences between the mean student-teacher ratio across groups or School Types were minimal, ranging from 21.68 to 22.09 students per teacher, and the overall student-teacher ratio for the district was 21.86 (Table 1). In the comparison of means for student-teacher ratios, there were no statistically significant differences ($p > .05$, $\eta^2 = .001$) between the three School Types. Similarly, there were no significant differences in in the variance pertaining to the availability of in-school or affiliated early childhood education programs across School Types ($p > .05$, $\eta^2 = .065$).

The individual time allotments in minutes that educators reported in the *Belize Educator Survey* were aggregated and reported by school type and overall for both English and Mathematics (Table 2). Mean comparisons indicated that the differences across School Types by subject areas were not statistically significant, English ($p > .05$, $\eta^2 = .154$) and Mathematics ($p > .05$, $\eta^2 = .051$). The average time allotment reported for English (82.5 minutes) was greater than the time allotment for Mathematics (66.4 minutes). Across School Types, the difference between highest and lowest average time allotted for English was 3.98 minutes, and for Mathematics it was 1.58 minutes (Table 2). The lack of variances in the student-teacher ratios, the access to early-childhood education, and the time allotted for English and Mathematics limited the meaningfulness of comparisons with either as an independent variable.

Table 2

Educator-Reported Time Investment (Minutes) Spent on English and Mathematics

School Type and Subject	Time in Minutes Min	Max	*M*
English			
Catholic	25.0	180.0	83.71
Non-parochial*	25.0	180.0	80.11
Other Denominations**	45.0	120.0	84.09
Mathematics			
Catholic	25.0	120.0	67.18
Non-parochial*	25.0	95.0	66.09
Other Denominations**	35.0	120.0	65.60
Overall School Level			
English	25.0	180.0	82.54
Mathematics	25.0	120.0	66.39

* Includes schools managed either by the Belize Ministry of Education or community-based organizations (not church-affiliated)

** Includes schools managed by leadership in the following churches: Seventh Day Adventist, Anglican, Assembly of God, Baptist, Church of Christ, and Methodist

Estimation

The models in this study stemmed from data collected with the *Belize Educational Survey* and from the Belize Ministry of Education. The initial descriptive statistics directly informed the formulation of the variables of the Base-Model. Thereafter, the Base-Model was used as the foundation for both estimation models, Weighted-Indicator Scores Protocol (WISP) and Multilevel Achievement Estimation Protocol (MAEP). Subsequently in this section, the models were described using statistics from analyses associated the protocols. For the Base-Model, the data collected from the educators were analyzed to confirm composition and determine values of the estimation variables. As the foundation for the two models of interest in this study, the Base-Model variables were processed according to associated protocols. This section presents results of significant steps to determine the final values for the variables for the district and schools.

Base-Model Statistics

Estimation variable scores were derived from data collected in the *Belize Educator Survey* and from data provided by the Belize Ministry of Education. The estimation variables in this study were School Context, Educator Efficacy, Opportunity to Learn, Educators' Perceptions of Their School Environment, School-Based Averages, School-Rigor, and the 2018 Primary School Examinations results. For scores based on the *Belize Educator Survey*, the School-Based Averages variable (*n* = 4) had the least number of items or components (Table 3). School-Rigor, which only had two components, was a school-level variable that was a derivative of School-Based Averages and Primary School Examinations.

Table 3

Descriptive Statistics for Overall Base-Model Variables (Score Range 0–500)

Indicator	*n* Components	Sample size	*Median*	M	*SD*	*SEM*
CTX	10	294	196.12	204.41	49.37	2.88
EFF	20	294	388.89	384.21	54.68	3.19
OTL	23	294	324.81	325.33	25.72	1.50
PRC	29	294	385.06	375.72	65.36	3.81
SBA	4	294	397.70	395.06	33.33	1.94
RIG	2	35	345.87	333.71	74.93	12.66
PSE	4	35	282.09	275.49	40.38	6.83

Note. Estimation variables: School Context (CTX), Educator Efficacy (EFF), Opportunity to Learn (OTL), Educators' Perceptions of Their School Environment (PRC), School-Based Averages (SBA), School-Rigor (RIG), Primary School Examinations (PSE)

For comparability, the score range possible for each variable was 0–500. The differences between the medians and means for each variable were minimal and ranged from (*SD* = 0.020) to (*SD* = 0.168). The School-Based Averages variable had the highest sample mean (*M* = 395.06, *SD* = 33.33), and the School Context variable had the lowest sample mean (*M* = 204.41, *SD* = 49.37) (Table 3). Both the School-Based Averages and School Context variable were based on the sample (*n* = 294) from individual educator responses. The variables School-Rigor and Primary School Examinations were composed using school-level statistical components and based on the sample of schools (*n* = 35). The sample means ranged from (*M* = 371.27) with (*SD* = 42.27) to (*M* = 275.49) with (*SD* = 40.38) for school-level variables. The first product resulting from the data collected from the Educators and the Belize Ministry of Education was the Base-Model. Essentially, the Base-Model comprised five overarching estimation variables and more specific variables representing key areas of interest used to measure and evaluate School-Level Achievement in this study.

Weighted-Indicator Scores Protocol (WISP)

The variables that composed the Base-Model for this study became the foundation for Weighted-Indicator Scores Protocol (WISP). Estimating School-Level Achievement using the WISP approach involved three main steps: conducting the data checks, determining the weights for the variables, and estimating achievement. Analyses for the WISP were conducted using SPSS version 24 (IBM Corporation, 2016).

Data Check

In terms of representativeness of the Base-Model data set, there were respondents from 36 of 38 (94.7%) public schools who participated in the *Belize Educator Survey*. After a closer look at the data set, it was determined that given the analyses included in the WISP, three schools had questionable representation. Two schools had insufficient representation (fewer than four educator responses), and one school without an eighth-grade class had no scores for the Primary School Examinations. Item-level assessment of the data quality included analyses for internal consistency, normality of distribution, homogeneity of variances, and integrity. Although these data checks were included in the WISP, the data checks were relevant to confirm or negate viability (quality) of estimation variables subsequently used in both models.

Internal Consistency

Assessment of the internal consistency quantified the reliability (coefficient alpha) to determine whether the items composing the measure are internally consistent. The internal consistency statistics are presented in Table 4. The coefficient alphas for the estimation variables ranged from $\alpha = .734$ to $\alpha = .956$ across variables. The variable with significant imputation, School-Based Averages with $\alpha = .763$, was retained in the model. The three most reliable variables were the 2018 Primary School Examinations ($\alpha = .956$), based on four items; Educators' Perceptions of Their School Environment ($\alpha = .934$), derived from 29 items; and Educator Efficacy ($\alpha = .927$), derived from 20 items. Based on the coefficient alphas ($\alpha > .7$), the estimation variables were deemed more than adequate for use in this study in which there are no intended highstakes decisions for the schools involved. Therefore, the variables remained intact with all their designated items. The conceivably strong coefficient alphas (reliability indicator $\alpha > .8$) increases confidence that the measures are likely to measure what they are intended to measure. However, in future applications, revision is recommended to increase reliability for the variables for the variables with $\alpha < .8$ (Opportunity to Learn, School Context, and School-Based Averages).

Table 4

Internal Consistency (Reliability) Statistics for Estimation Variables

Estimation Variables	Items	Coefficient Alpha α
School Context (CTX)	10	.749
Educator Efficacy (EFF)	20	.927
Opportunity to Learn (OTL)	23	.734
Educators' Perceptions of Their School Environment (PRC)	29	.934
School-Based Average (SBA)	4	.761
Primary School Examinations (PSE)	4	.956

Shapiro-Wilk and Levene

The Shapiro-Wilk test for normality verified that the variables in the Base-Model did not violate the assumption of normality (Table 5). The *p*-values for the educator-level estimation variables, School Context, Educator Efficacy, Opportunity to Learn, Educators' Perceptions of Their School Environment, and School-Based Averages, were at very satisfactory values ($p < .001$).

Levene's test for homogeneity of variance tested the null hypothesis that the variance was equal across groups (Table 5). Therefore, statistical significance indicated a violation of the assumption of independence. All analyses were not statistically significant. The results from the Shapiro-Wilk and Levene tests indicated that the data were of appropriate quality for analysis of variance (ANOVA) and other procedures.

Table 5

Normality and Homogeneity Assumptions

Estimation Variables	Shapiro-Wilk			Levene			
	statistic	*df*	*P*	statistic	*df1*	*df2*	*p*
CTX	.961	290	.000	1.319	2	32	.281
EFF	.971	290	.000	.218	2	32	.805
OTL	.932	290	.000	2.114	2	32	.137
PRC	.957	290	.000	.215	2	32	.807
SBA	.971	290	.000	.671	2	32	.518

Note. Estimation Variables: School Context (CTX), Educator Efficacy (EFF), Opportunity to Learn (OTL), Educators' Perceptions of Their School Environment (PRC), School-Based Averages (SBA)

One-Way ANOVA

The one-way analysis of variance (ANOVA) procedure tested for differences in the estimation variables, using the 2018 Primary School Examination scores as the independent variable, to determine whether there were significant statistical differences. The one-way ANOVA at the ($\alpha = .05$) level confirmed statistically significant differences $F(34, 255) = 31.60$ ($p < .001$) to-$F(34, 255) = 1.77$ ($p < .01$) between Primary School Examinations and each of the estimation variables (Table 6). The *F*-test statistics confirmed with reasonable confidence that there were statistically significant variances in the five primary estimation variables of the data sample. Furthermore, the significant results indicated that the relationships between the estimation variables and Primary School Examinations were not due to random sampling or chance. Therefore, the important takeaway was that Base-Model data were of viable quality for the other analyses included in the WISP.

Table 6

One-Way ANOVA

		M	*df*	Mean Square	*F*	*p*	η^2
CTX	Between Schools	184.15	34	16740.91	31.60	.000	.808
	Within School	111.03	255	529.78			
	Total		289				
EFF	Between Schools	385.70	34	4869.82	1.77	.007	.191
	Within School	383.56	255	2747.89			
	Total		289				
OTL	Between Schools	348.23	34	2065.83	4.44	.000	.372
	Within School	324.87	255	465.57			
	Total		289				
PRC	Between Schools	376.06	34	13623.55	4.49	.000	.374
	Within School	374.96	255	3036.31			
	Total		289				
SBA	Between Schools	403.90	34	1885.38	1.95	.002	.206
	Within School	394.23	255	967.01			
	Total		289				

Note. Listwise n = 35.

Dependent: School Context (CTX), Educator Efficacy (EFF), Opportunity to Learn (OTL), Educators' Perceptions of their School Environment (PRC), School-Based Averages (SBA)

Independent: Primary School Examinations (PSE)

Correlations

Level-One Comparisons (n = 290 Educators)

Six relationships between the variables in level one, or the overall district by educators (n = 290), were flagged as statistically significant (Table 7). Pearson correlations indicated a statistically significant positive relationship between School Context and Opportunity to Learn (r = .134, p < .05). Pearson correlations indicated statistically significant positive associations between Educator Efficacy and Opportunity to Learn ($r = .261$, $p < .001$). There was also a positive statistically significant relationship between Educator Efficacy and the Educators' Perceptions of Their School Environment ($r =.506$, $p < .001$) and between Educator Efficacy and School-Based Averages ($r = .207$, $p < .001$). Similarly, Pearson correlations indicated statistically significant positive associations between Educators' Perceptions of Their School Environment and Opportunity to Learn ($r = .338$, $p < .01$) and between Educators' Perceptions of Their School Environment and School-Based Averages ($r = .188$, $p < .001$).

Level-Two Comparisons (n = 35)

Pearson correlations (Table 7) indicated that there were statistically significant relationships between the five overall estimation variables and the two school-level variables assessed at the school level ($n = 35$ schools). The two school-level variables included were Primary School Examinations and School-Rigor. Among the school-level comparisons, there were seven significant relationships flagged as statistically significant. At the school level, Pearson correlations did not indicate any statistically significant association between the School Context variable and the other variables (Table 7).

Pearson correlations indicated statistically significant positive associations between Educator Efficacy and Opportunity to Learn ($r = .422$, $p < .05$) and between Educator Efficacy and Educators' Perceptions of Their School Environment ($r = .483$, $p < .01$). Positive statistically significant correlations were also indicated for associations between Opportunity to Learn and Educators' Perceptions of Their School Environment ($r = .562$, $p < .001$), between Opportunity to Learn and Primary School Examinations ($r = .466$, $p < .01$), and between Opportunity to Learn and School-Rigor ($r = .416$, $p < .05$).

Additionally, Educators' Perceptions of Their School Environment also correlated statistically significantly with School-Based Averages (r = .413, p < .05). Since School-Rigor was computed based on Primary School Examinations, the association between the two variables correlated with strong statistical significance (r = .991, p < .001).

Table 7
Correlations between Estimation Variables at the School Level

	M	*SD*	CTX	EFF	OTL	PRC	SBA	PSE	RIG
Level One (educators)									
CTX	205.06	49.37	1	-.022	.134*	-.054	-.110	.316**	.447**
EFF	383.56	54.75		1	.261**	.506**	.207**	.040	.007
OTL	324.87	25.57			1	.338**	.115	.128*	.078
PRC	374.96	65.44				1	.188**	.117*	-.011
SBA	394.23	32.79					1	.101	-.085
PSE	279.30	39.53						1	.889**
RIG	342.78	73.12							1
Level Two (schools)									
CTX	184.15	38.27	1	-.147	-.164	-.159	-.303	.150	.189
EFF	385.70	25.60		1	.422*	.483**	.133	.184	.168
OTL	348.23	21.39			1	.562**	.260	.446**	.416*
PRC	376.06	39.94				1	.413*	.253	.199
SBA	403.90	19.17					1	.113	-.016
PSE	275.49	40.38						1	.991**
RIG	333.71	74.93							1

* Correlation is significant at the .05 level (2-tailed).

** Correlation is significant at the .01 level (2-tailed).

Estimation Variables: School Context (CTX), Educator Efficacy (EFF), Opportunity to Learn (OTL), Educators' Perceptions of Their School Environment (PRC), School-Based Averages (SBA), School-Rigor (RIG), Primary School Examinations (PSE)

Level One listwise n = 290, df = 289

Level Two listwise n=35, df = 34

Confirmatory Factor Analyses (CFA)

To further evaluate the psychometric properties of the estimation variables, the Kaiser-Meyer-Olkin Test (KMO) and Bartlett's Test of Sphericity (Table 8) were included in the confirmatory factor analyses (CFA) for each variable. The CFAs used the Maximum Likelihood (ML) extraction method on educator-level (n = 294) Base-Model data and conducted for each variable independently.

KMO and Bartlett's

All estimation variables had satisfactory values (p > .50) on the KMO test. The estimation variables with the highest values were Educator Efficacy and Educators' Perceptions of Their School Environment, with KMO test statistics of (p = .937) and (p = .935), respectively (Table 8). The Opportunity to Learn variable had the most substantial proportion of variance explained (73.18%) by the Kaiser's rule, but the lowest (p = .530) KMO test statistic. As reported in Table 8, the Bartlett's test statistics were less than .001 for all the estimation variables in the data set. The low test statistic indicated that the items that comprised the estimation variables presented a cohesive structure.

Table 8

Kaiser-Meyer-Olkin Test (KMO) and Bartlett's Test of Sphericity

Estimation Variable	Items	KMO Test	Bartlett's Test		
			Chi-Square	*df*	*p*
CTX	10	.77	4462.51	45	.00
EFF	20	.94	3353.47	190	.00
OTL	23	.53	411.27	253	.00
PRC	29	.93	5653.98	406	.00
SBA	4	.75	321.36	6	.00
PSE	4	.86	148.45	6	.00

Estimation Variables: School Context (CTX), Educator Efficacy (EFF), Opportunity to Learn (OTL), Educators' Perceptions of Their School Environment (PRC), School-Based Averages (SBA), Primary School Examinations (PSE)

Confirmatory Factor Analyses (CFA) Outcomes

Apart from confirming the dimensionality of the items composing the estimation variables, the CFAs on the variables were used as a measure of variance explained by the extracted factors were used in the determination of the weights applied to variables in the Weighted-Indicator Scores Protocol. In each confirmatory factor analysis with maximum-likelihood extraction and varimax rotation (Table 9), the Eigenvalues and percentage of variance accounted for by the first factor extracted exceeded 20% of the total variance explained (Table 9). The variance explained by the first factors across the five estimation variables and Primary School Examinations ranged from 24.21% for Opportunity to Learn to 88.71% for Primary School Examinations (Table 9).

Using Kaiser’s rule, there were three interpretable factors for School Context. The first factor in School Context accounted for 41.2% of the variance. Cumulatively, the three factors recognized with Kaiser’s rule explained 69.92% of the variance in the estimation variable for School Context (Table 9). An examination of the scree plot (Cattell’s scree-test) for the School Context variable illustrated that in School Context, three factors were greater than one. There was also a very sharp decline (cliff-like drop) between the first and second factors that indicated the distinctive nature of the first factor as a primary factor or focus in the construct for School Context (Figure 3). After the drop between the second and third factors extracted, the plot leveled out (the way scree or fallen debris on the mountainside would appear after it becomes stable). There were 10 items in this variable. Based on observation of the scree plot, two factors were most prominent.

Four factors of Educator Efficacy had Eigenvalues greater than one. The first factor accounted, for 44.35% of the variance in the variable (Table 9). Cumulatively, the four recognized factors explained 62.37% of the variance in the construct for Educator Efficacy. The scree plot reflected a sharp difference between the first and second factor (Figure 3). Beyond the very sharp difference between the first and second factors extracted, the slope leveled off. From observation of the scree plot, two factors were above the baseline, but one factor was the most prominent. This observation was consistent with focus of the 20 items that drew attention to the educator with the format of the question stem. A very similar pattern was observed for the Educators' Perceptions of Their School Environment estimation variable that had 29 items. Five factors were identified using Kaiser's rule. The first factor accounted for 41.88% of the variance in Educators' Perceptions of Their School Environment, and cumulatively, the five recognized factors explained 61.86% of the variance (Table 9). Like the observations for the Educator Efficacy variable, the scree plot for the Educators' Perceptions of Their School Environment variable also had a high value for the first factor (12.15) and a steep decline to the second factor (2.06), after which the scree leveled of considerably by the third and consecutive factors.

Seven factors with Eigenvalues greater than one were evident for the Opportunity to Learn variable, the most factors extracted of all estimation variables. The first factor accounted for 24.21% of the variance in the construct (the lowest percentage of all first factors). Cumulatively, the seven factors explained 73.18% of variance in this estimation variable (Table 9). Subsequent analysis, the scree-test, showed steep declines between the first thee variables and the leveling out began after the sixth factor. The Opportunity to

Learn indicator comprised 23 items covering the general construct from multiple angles (foci) and the trend in the results appears consistent with.

The two academic related variables, School-Based Averages and Primary School Examinations had equal number of items (4) and only one factor extracted using Kaiser's rule (Table 9). For the School-Based Averages estimation variable, the first factor accounted for 58.9% of the variance and for the Primary School Examinations, the first factor accounted for 88.7% of the variance. Observations from the corresponding scree plots support the unidimensionality of these two variables (Figure 4). Unidimensional determinations (only one factor extracted) were revealed for both the School-Based Averages and the Primary School Examinations variables for which the same four subject areas informed academic achievement. The prominence of each of the first factors extracted in each estimation variable and the number of factors extracted were also indications of how narrowly focused the items were to the corresponding constructs being measured (beyond subject dominance).

Table 9

Confirmatory Factor Analysis

		1st Factor extracted		Factors extracted with Eigenvalues greater than one	
Base-Model Variables	Items	Initial Eigenvalue*	% variance explained	# Factors extracted	Total variance explained (%)
CTX	10	4.12	41.20	3	69.92
EFF	20	8.87	44.35	4	62.37
OTL	23	5.57	24.21	7	73.18
PRC	29	12.15	41.88	5	61.86
SBA	4	2.34	58.51	1	58.51
PSE	4	3.55	88.71	1	88.71

* Extraction Method = Maximum Likelihood (ML) with varimax rotation

School Context (CTX), Educator Efficacy (EFF), Opportunity to Learn (OTL), Educators' Perceptions of Their School Environment (PRC), School-Based Averages (SBA), Primary School Examinations (PSE)

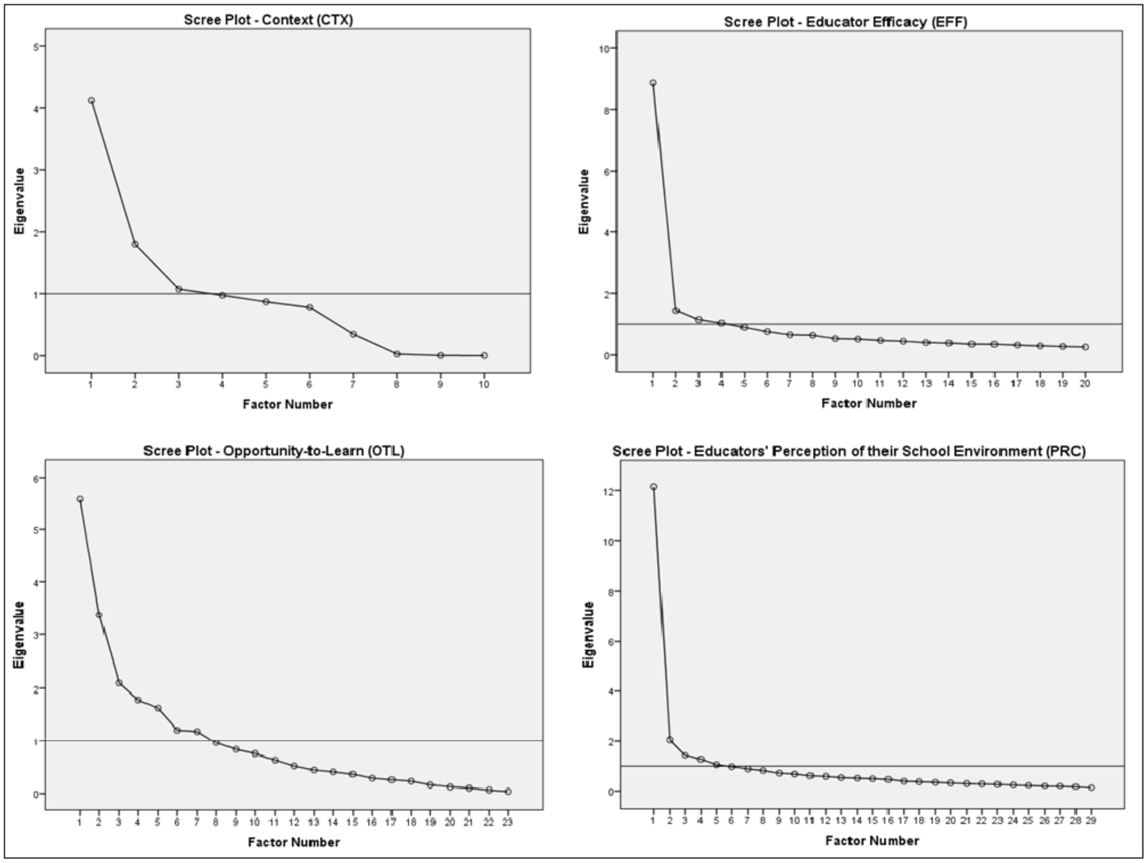


Figure 3. Scree Plots for the nonacademic variables: Context (CTX), Educator Efficacy, Opportunity to Learn (OTL), Educators' Perceptions of Their School Environment (PRC)

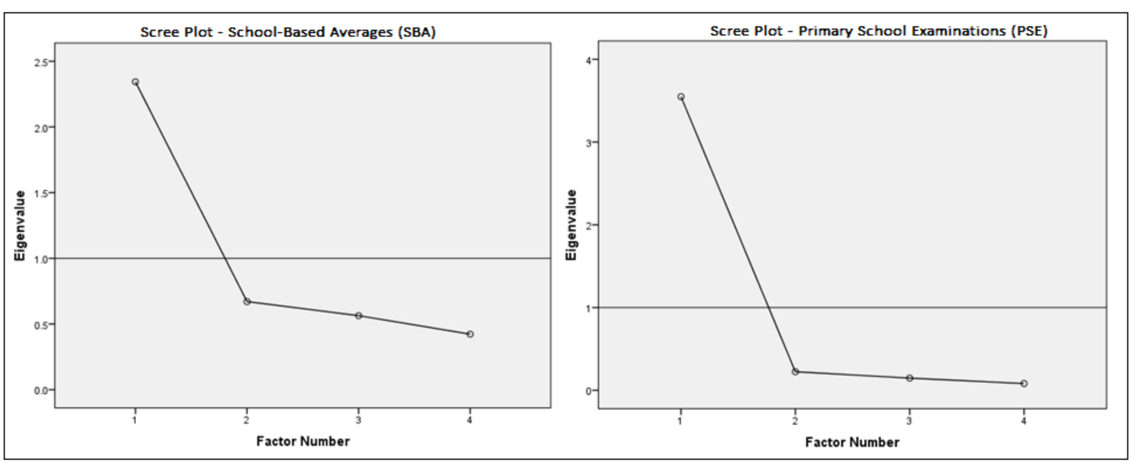


Figure 4. Academic variables analyzed: School-Based Averages (SBA) and Primary School Examinations (PSE)

Weighted-Indicator Scores Protocol (WISP) Achievement Estimation

The weights were applied to the Base-Model scores to generate the school-level scores for Weighted-Indicator Scores Protocol estimation variables. The relationships between variables were reassessed to confirm whether or not there were changes the relationships between the variables. The Pearson correlation values for the WISP variables matched precisely with those previously reported for the Base-Model variables in Table 7. The WISP transformed the variables (weighted), but the transformation did not change the fundamental relationships between estimation variables. Therefore, any subsequent negative or positive differences noted between estimates derived from the Base-Model and those derived from the WISP were attributable to the transformation (weights).

The overall descriptive statistics for estimates (school district and five school-level examples) derived using the WISP are reported in Table 10. Among schools represented in the district (n = 35), the unstandardized weighted score means, on the 0 to 500 gauge, ranged from (M = 142.93) for School Context variable to (M = 239.63) for Educator Efficacy. However, unlike the Base-Model values, which were based on an equally possible score range of 0–500 points, these estimates were weighted accordingly with the amount of variance explained. Therefore, a meaningful comparison of differences using the unstandardized estimates must appropriately be limited to within-variable comparisons. Estimates for four schools were included in Table 10 to illustrate general profile comparisons to the district mean. For example, the Educator Efficacy statistics for the district were (M = 239.63, SD = 34.10). On average, self-reports for Educator Efficacy in School-01 (M = 197.70) was 1.2 standard deviations less than the district mean. However, the mean for Educator Efficacy in School-02 (*M* = 259.88) was .59 standard deviation above the district mean. The difference between School-01 and School-02 was 1.79 standard deviations. Therefore, the differences in Educator Efficacy compared to the district mean and between these schools (examples) were not statistically significant (critical difference threshold defined as greater than or equal to 1.96 standard deviations to indicate 95% confidence level or statistical significance). Considering the School-Rigor variable, the statistics highlighted include School-01 (*M* = 346.78), School-31 (*M* = 168.13), and the district (*M* = 333.71, *SD* = 74.93). The difference between School-01 and the district showed that the School-Rigor estimated for School-01 was .17 standard deviation higher than the district mean. The difference between School-31 and the district showed that the School-Rigor estimated for School-31 was 2.21 standard deviations less than the mean for the school. Referencing the critical difference threshold, the difference between School-01 and the district was not statistically significant. However, the differences between School-31 and the district mean and School-01 were greater than 1.96 standard deviations, and as such, those differences in the School-Rigor estimates were statistically significant.

The standardized WISP estimates for schools (*n*=35) had an overall mean and standard deviation of *M* = 0 and *SD* = 1 (*SE* = .169). Given the examples presented, there were statistically significant differences between School-31, the district, and the other three schools. The differences in means exceeded 1.96 standard deviations in each pairwise comparison (Table 10). The mean differences between the district and other schools, and also between the school examples in pairwise comparisons, were each less than 1.96 standard deviations, thus were not statistically significant. The full complement of standardized School-Level Achievement statistics estimated using the WISP described is reported in Appendix F.

Table 10

The District-Level and Five Selected School-Level Examples of WISP Variable Estimates

	Unstandardized Weighted-Indicator Scores Protocol					Variable	Achievement Estimates	
	CTX	EFF	OTL	PRC	SBA	RIG	WISP	WISP Std.
District-level								
M	142.93	239.63	238.08	232.42	231.15	333.71	237.80	0
SD	34.52	34.10	18.82	40.43	19.50	74.93	17.60	1.00
SEM	2.01	1.99	1.10	2.36	1.14	12.66	2.97	.169
School-level examples								
School-01	121.77	197.70	253.93	246.49	256.05	346.78	237.12	-.04
School-02	136.38	259.88	259.98	253.60	235.60	392.33	256.29	1.05
School-27	190.90	242.42	280.59	260.85	227.92	398.90	266.93	1.65
School-31	142.82	220.61	234.98	205.41	231.99	168.13	200.66	-2.11

Unstandardized Weighted-Indicator Scores Protocol: School Context (CTX), Educator Efficacy (EFF), Opportunity to Learn (OTL), Educators' Perceptions of Their School Environment (PRC), School-Based Averages (SBA) Unstandardized school-level variable: School-Rigor (RIG)
Unstandardized achievement estimate: Weighted-Indicator Scores Protocol (WISP)
Standardized achievement estimate: Weighted-Indicator Scores Protocol (WISP Std.)
Total number of schools included in the analyses (*n*=35)

Multilevel Achievement Estimation Protocol (MAEP)

Multilevel Achievement Estimation Protocol (MAEP) was estimated within a multilevel structural equation model framework for two levels: the educator level (level one) and the school level (level two). Multilevel structural equation model analyses were conducted using M*plus*® version 8.3 (Muthen & Muthen, 1998–2019). The analyses used a maximum likelihood estimator with robust standard errors (MLR) and the numerical integrations algorithm set to 7 integrations per dimension.

Procedurally, the multilevel structural equation model input (Figure 2, p. 91) and estimations (Appendix E) terminated normally and without structural warnings after two schools with insufficient variability (less than four participants) were excluded. The multilevel structural equation model procedures included 34 school clusters. In the multilevel analysis, there was one missing data pattern for the school. That missing pattern was for the same school previously excluded from WISP analyses because there were no associated Primary School Examinations scores.

Descriptive Statistics

Comparatively, the descriptive statistics in level one (educator level) of the multilevel structural equation model for the Multilevel Achievement Estimation Protocol were consistent with the Base-Model statistics previously reported in Table 1, thus confirming that processing the data with different software did not change the inherent structure of the data.

The level two (school clusters) model results are reported in Table 11. The intercepts in the multilevel structural equation modeling analyses were School Context, Educator Efficacy, Opportunity to Learn, Educators' Perceptions of Their School Environment, and School-Based Averages. The analyses revealed that estimates for the intercepts contributed to the model with statistical significance ($p < .001$); intercept means ranged from $M = 203.06$ to $M = 387.32$. The random slope, Educator Effect/Bias, introduced in this model showed nonsignificant, negative estimates ($M = -13.30$, $p > .05$). Similarly, statistically nonsignificant results were revealed in all the regression analyses in the model: Educator Effect/Bias regressed on the level-two Achievement Between Schools (L2-ABS) (R^2 = 5.19, p > .05) and Educator Effect/Bias regressed on School-Rigor (R^2 = .10, p > .05) (Table 11).

Table 11

*Multilevel Structural Equation Modeling Results (M*plus®*)*

Intercepts	Mean	*SE*	Est./SE	Two-Tailed *p*-value
CTX	203.06	10.55	19.25	.001
EFF	387.32	4.11	94.20	.001
OTL	329.18	6.08	54.13	.001
PRC	38.35	13.31	28.58	.001
SBA	379.65	11.46	33.14	.001
EEB (slope)	-13.30	76.69	-0.17	.862
Regression	***Estimate***	***SE***	**Est./SE**	**Two-tailed *p*-value**
EEB ON L2-ABS achievement	5.19	51.90	0.10	.920
EEB ON RIG	0.10	0.25	0.38	.702
SBA ON L2-ABS	45.07	23.06	1.95	.051
SBA ON RIG	0.05	0.03	1.62	.105

Variables: School Context (CTX), Educator Efficacy (EFF), Opportunity to Learn (OTL), Educators' Perceptions of Their School Environment (PRC), School-Based Averages (SBA), Educator-Effect/Bias (EEB), School-Rigor (RIG), Level 2 Achievement Between Schools (L2-ABS)

The Multilevel Achievement Estimation Protocol (MAEP) values were computed based on the results of the multilevel structural equation modeling analyses. In the process of estimating the descriptive, Pearson correlations showed that the relationships between school-level variables were identical to those previously reported correlation statistics reported for the Base-Model (Table 7). In effect, the method for computing final MAEP achievement estimates did not affect the relationships (structure) between the variables.

MAEP estimates are reported in Table 12 for the district level and the four schools previously used as examples. The full complement of MAEP achievement estimates by school are reported in Appendix F. At the district level, the general descriptive statistics (means, standard deviations, and standard errors of the mean) for the variables, the overall achievement estimate, and the standardized version of the achievement estimate are presented. The estimates presented for schools were intended to illustrate some possible school profile comparisons.

Of the four school-level examples (Table 12), School-01 had the lowest score for School Context (CTX = 16.25), a difference of 3.12 or 0.90 standard deviation below the district mean. The difference between the School Context estimate for School-01 and the district was not statistically significant ($p > .05$). Of these four schools, School-27 had the highest estimate for School Context (CTX = 25.88), a difference of 6.51 or 1.88 standard deviations above the district mean. Even though the estimate was above the district mean, that difference was not statistically significant ($p > .05$). However, the difference between the School Context estimates for School-01 and School-27 (9.63 or 2.78 standard deviations) was statistically significant ($p < .05$).

The four schools included for comparison did not differ from the district mean or between schools with statistical significance in reference to the School-Based Averages ($p > .05$). The differences were less than or equal to 1.71 or 0.59 standard deviation (Table 12). Given the overall School-Level Achievement estimates derived from the Multilevel Achievement Estimation Protocol in the standardized format (MAEP Std.), School-27 had the highest estimate (MAEP = 1.442), and School-31 had the lowest estimate (MAEP = 207.20) of the four schools. The district mean was (MAEP = -.052);

therefore, the differences were not statistically significant (*p* > .05) between the district and School-27 (1.49 or 1.5 standard deviations) and School-31 (-1.26 or 1.27 standard deviations). However, the difference between the standardized Multilevel Achievement Estimation Protocol estimates for School-27 and School-31 was statistically significant (p < .05), 2.75 or 2.76 standard deviations. The School-Level Achievement estimates for both schools (27 and 31) were consistent within the expected or normal district range, but with both on opposite sides of the mean, the cumulative difference exceeded the critical difference within 1.96 standard deviation or normal range (95%).

Table 12

Multilevel Achievement Estimation Protocol (MAEP) Estimates

						Overall Achievement Estimates	
District Level	CTX	EFF	OTL	PRC	SBA	MAEP	MAEP Std.
Mean	19.37	93.44	53.50	28.23	34.48	221.46	0
SD	3.47	13.30	4.23	4.91	2.91	10.85	1.00
SE	0.27	0.78	0.25	0.29	0.17	1.81	0.171
School-level Examples							
Schoo-01	16.25	77.39	52.80	29.60	35.71	208.20	-1.221
School-02	18.49	101.33	58.42	30.80	35.14	232.39	1.007
School-27	25.88	94.52	63.05	31.68	34.00	237.11	1.442
School-31	19.36	86.02	52.80	24.95	34.60	207.20	-1.314

Variables: School Context (CTX), Educator Efficacy (EFF), Opportunity to Learn (OTL), Educators’ Perceptions of Their School Environment (PRC), School-Based Averages (SBA)
Unstandardized achievement estimate: Multilevel Achievement Estimation Protocol (MAEP)
Standardized achievement estimate: Multilevel Achievement Estimation Protocol (MAEP Std.)

Research Questions

There were five questions asked in this study. The findings from the analyses used to answer the questions are presented in this section.

Question 1

Is there a difference in the overall School-Level Achievement estimates using a classic summative approach with the Weighted-Indicator Scores Protocol (WISP) versus a latent approach with a Multilevel Achievement Estimation Protocol (MAEP)?

Multiple statistical analyses were conducted to determine whether there were statistically significant differences between the school-level estimates derived from Weighted-Indicator Scores Protocol (WISP) and from Multilevel Achievement Estimation Protocol (MAEP). Model estimates were standardized ($M = 0$, $SD = 1$) for the meaningful comparability. With a common mean and standard deviation, the descriptive statistics reported (Table 13) included the number of schools with estimated values, standard error of the means, score range, minimum, and maximum values. The standardized estimates were reported for Primary School Examinations, and the three estimation models: the Base-Model, WISP, and MAEP. Using listwise deletion, 33 of the 36 schools represented were retained for comparison.

The standard errors of the means were fairly consistent for the standardized Primary School Examinations averages and the model estimates. Except for MAEP ($SEM = 0.171$), the standard errors of the mean for the models ($SEM = .169$) were the same across models (Table 13).

The widest range in scores (distance across distribution) across models was revealed for MAEP (4.348) and WISP (4.182). The range of scores for the Base-Model (4.026) and Primary School Examinations (4.052). However, the largest difference in the value for observed score ranges, between Primary School Examinations and Multilevel Achievement Estimation Protocol (0.296), was not statistically significant considering differences in terms of standard deviations (*SD* < 1.96). Similarly for the question at hand, whether or not the estimates derived from WISP and MAEP were different, considering the bandwidths of the score ranges, the difference (0.192) between the two models was not statistically significant (*SD* < 1.96).

Table 13

Score Range for Standardized PSE Averages and Model Estimates

	n	*SEM*	Score Range	minimum	maximum
PSE	35	0.169	4.052	-2.487	1.565
BASE	35	0.169	4.026	-2.274	1.753
WISP	35	0.169	4.156	-2.501	1.655
MAEP	34	0.171	4.348	-2.258	2.091

Primary School Examinations (PSE), the Base-Model (BASE), Weighted-Indicator Scores Protocol (WISP), and Multilevel Achievement Estimation Protocol (MAEP)

Listwise deletion (*n* = 33)

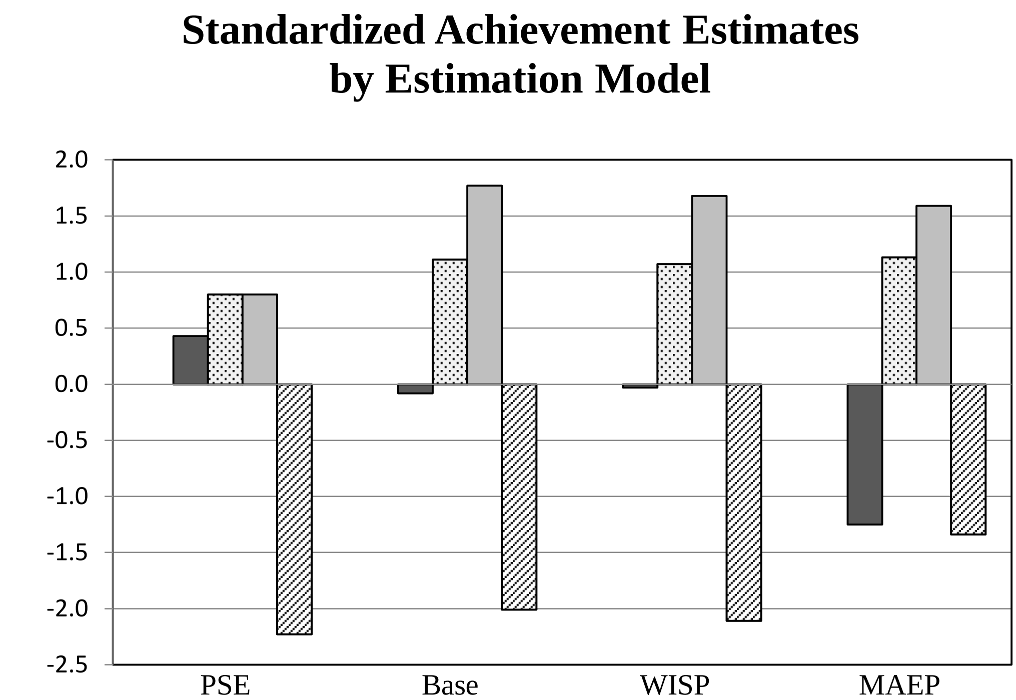
Rank-Order Disclaimer

The School-Level Achievement estimates in this study are not intended as a means of ranking schools and are not reported as such. Rank-ordering was considered only as a means of comparing the methodologies used in this study, not school quality. (The Belize Ministry of Education explicitly discourages use of the Primary School Examinations to rank schools.) The rank-ordered estimates for Primary School Examinations were matched to the rank-ordered school-level estimates for the Base-Model, the Weighted-Indicator Score Model (WISP), and the Multilevel Achievement Estimation Protocol (MAEP). For rank-ordered matching, WISP matched with Primary School Examinations either exactly or within two ranks for a total of 16 times out of 33 (48.5%). The MAEP rank-ordered estimates matched with Primary School Examinations either exactly or only off by one or two ranks 5 of 33 (15.2%). The Base-Model rank-ordered estimates matched with Primary School Examinations exactly or within two ranks 11 times out of 33 (33.3%). Although the Base-Model was the initial data source for WISP and MAEP, the WISP matched the Primary School Examinations 14.3% more than the Base-Model, and MAEP matched to the Primary School Examinations 17.1% less than the Base-Model. There were 12 of 33 matches (36.4%) between the WISP and MAEP rank-ordered estimates. Therefore, the highest percentage rank-order matching to Primary School Examinations involved estimates derived from the WISP.

The standardized School-Level Achievement estimates by model ($M = 0$, $SD = 1$), for the district and the four schools previously used as examples, are reported in Figure 5. The general pattern in the observed values of the School-Level Achievement estimates across models (BASE, WISP, and MAEP) was similar to the pattern for the schools' averages on the Primary School Examinations. Subsequent comparisons were made based on the logic that statistical significance expressed in differences between model estimates was achieved if the difference itself was greater than or equal to 1.96 standard deviations ($SD \geq 1.96$). School-27 consistently had the highest estimates in comparison to the district and other schools. However, the differences in the estimates with School-27 ranged from 0.80 to 1.77 standard deviations and were therefore not statistically significant. School-31 had the lowest estimates across each measure. The differences between estimates for School-31 and the estimates for district were statistically significant ($SD > 1.96$) for Primary School Examinations, the Base-Model, and WISP. Within the example, School-01 was consistently the lowest (Figure 5). However, the difference between School-31 and the district for MAEP was not statistically significant

(*SD* < 1.96). In comparison to the other schools, the observed pattern was similar between School-31, School-02, and School-27 (*SD* > 1.96). For the comparison between School-31 and School-01, the differences were mixed. For estimates derived from the Primary School Examinations and the WISP, the differences were statistically significant (*SD* > 1.96), but the differences in the estimates derived from the Base-Model and the MAEP were not statistically significant (*SD* < 1.96).

The across-model differences in estimates derived by each protocol (within the same school) were also compared using standard deviations as the metric (significant difference with $SD \geq 1.96$). Unlike the across-school comparisons where there were multiple instances of statistically significant differences (*SD* > 1.96), the across-model comparisons did not reveal any instances of a statistically significant difference. Differences across the models and within a school ranged from 0.02 to 1.68 standard deviations (*SD* < 1.96).



	PSE	Base	WISP	MAEP
District	0.00	0.00	0.00	0.00
School-01	0.43	-0.08	-0.03	-1.25
School-02	0.80	1.11	1.07	1.13
School-27	0.80	1.77	1.68	1.59
School-31	-2.23	-2.01	-2.11	-1.34

Figure 5. Model Comparison. The standardized Primary School Examinations (PSE) averages and School-Level Achievement estimates (n = 33 schools with listwise deletion) are by estimation model: Base-Model (Base), Weighted-Indicator Scores Protocol (WISP), and Multilevel Achievement Estimation Protocol (MAEP).

A *t*-test was used to further examine the differences between the standardized estimates for Weighted-Indicators Scores Protocol and Multilevel Achievement Estimation Protocol (Table 14). The *t*-test resulted in retaining the null hypothesis. The *t*-test revealed that there were no statistically significant differences between the standardized estimates for the WISP, t (34) = .000, $p > .05$, and the MAEP, t (33) = .064, $p > .05$.

Table 14

t-test

	Mean Difference	*t*	*df*	*p* (2-tailed)	95% Confidence Interval of the Difference	
					Lower	Upper
Weighted-Indicator Scores Protocol	0.000	0.000	34	1.000	-0.344	0.344
Multilevel Achievement Estimation Protocol	0.011	0.064	33	.949	-0.336	0.358

Correlations

The examination of relationships between estimation models, in terms of Pearson correlations, revealed that statistically significant relationships existed between all the models, Primary School Examinations, the Base-Model, Weighted-Indicator Scores Protocol (WISP), and Multilevel Achievement Estimation Protocol (MAEP). The relationship between Primary School Examinations and the Base-Model was (r = .829, p < .001), and between Primary School Examinations and WISP, the correlation was (r = .903, p < .001) (Table 15). The relationship between Primary School Examinations and MAEP (r = .354, p < .05) was also statistically significant, but at the .05 level. Similarly, the relationships between the Base-Model and the WISP (r = .696, p < .001) and with the MAEP (r = .792, p < .001) were also statistically significant at the .01 level (Table 15).

Table 15

Correlations between Standardized PSE and Model Estimates

	PSE	BASE	WISP	MAEP
PSE	1			
BASE	.829**	1		
WISP	.903**	.987**	1	
MAEP|	.354*	.792**	.696**	1

Primary School Examinations (PSE), the Base-Model (BASE), Weighted-Indicator Scores Protocol (WISP), and Multilevel Achievement Estimation Protocol (MAEP)

** Correlation is significant at the .01 level (1-tailed).

* Correlation is significant at the .05 level (1-tailed).

Listwise $n = 33$, $M = 0$, $SD = 1$

Multiple Regression Analysis

The final measure to determine if there was a difference between the two models, Weighted-Indicator Scores Protocol (WISP) and Multilevel Achievement Estimation Protocol (MAEP), was to identify which was the better predictor of Primary School Examinations. A multiple regression was run to predict Primary School Examinations with WISP and MAEP as the predictor variables.

The regression model met the assumptions for: continuous variables, linear distribution, no apparent outliers, independence of observation, homoscedasticity, and normally distributed residual errors. As previously established, the two predictor variables were significantly correlated. Therefore, there were low levels (within acceptable ranges) of autocorrelation (Durbin-Watson = 1.97) for the model. The negative value in the *Beta* statistics for MAEP (β = -.532, $p < .05$) was within reason since the multicollinearity statistic (VIF=1.939) was not at a level that warranted concern (Table 16).

It was determined that both WISP and MAEP predicted Primary School Examinations at a statistically significant level, ($F(2, 30) = 366.678$, $p < .001$, $R^2 = .961$). Together, both WISP and MAEP added significantly to the model ($p < .001$) and the adjusted R^2 indicating 95.8% of the variance accounted. Nevertheless, no significant differences were teased out to determine which of WISP and MAEP was the better predictor.

Table 16

Regression with WISP and MAEP as the Predictor Variables

Model Summary

		Adjusted		Change Statistics					Durbin-
R	*R^2*	*R^2*	*SE*	*R^2*	*F*	*df1*	*df2*	*p*	Watson
.980	.961	.958	.202	.961	366.678	2	30	.000	1.957

Coefficients

	Unstandardized		Standardized			
	B	*SE*	*Beta*	*t*	*p*	VIF
(constant)	-.046	.035		-1.300	.204	1.939
					.000	
WISP	1.266	.050	1.273*	25.257		
MAEP	-.560	.053	-.532*	-10.560	.000	

Dependent variable: Primary School Examinations (PSE)

Predictor variables: Weighted-Indicator Scores Protocol (WISP), Multilevel Achievement Estimation Protocol (MAEP)

* Statistically significant at the .05 level

Indications from the various measures of differences between the standardized estimates for WISP and MAEP did not yield either apparent or statistically significant results. Differences were evaluated using descriptive statistics (including score ranges and rank-order comparisons). Differences were also measured in terms of standard deviations, relationships described by Pearson correlations, a *t*-test, and a linear regression (predicting the 2018 Primary School Examinations). It was determined that the School-Level Achievement estimates derived from WISP and MAEP were not significantly different. Therefore, the answer is no to the research question about differences in the overall estimates for school-level educational achievement using a latent approach with a MAEP versus a classic summative approach with the WISP. The estimates produced by both methods were very similar and equally viable for estimating School-Level Achievement with the kind of sample assessed (a school district in Belize).

Question 2

Were there differences in the School-Level Achievement estimates by selected school characteristics such as School Type (Catholic, Non-parochial, or Other Denominations) or Location (Urban, Rural, or Remote)?

The analyses to answer the question were in two parts, first to determine if School Type impacted educational achievement with statistical significance (Table 17) and, second, to understand whether educational achievement was impacted by physical location (Table 18).

Using one-way analyses of variance tests (ANOVA) of school-level achievement, it was determined that the differences were not statistically significant for Catholic, Non-parochial, or Other Denominations, regardless of which mode of estimation was used. The School Type, or schools managed by Catholic, Non-parochial, or Other Denominations, did not appear to have statistically significant bearing on the School-Level Achievement estimates derived using Weighted-Indicator Scores Protocol (WISP) ($F(2, 30) = 1.198$, $p > .05$) Similarly, the differences in School-Level Achievement estimates derived from Multilevel Achievement Estimation Protocol (MAEP) were not statistically significant by School Type, or schools managed by Catholic, Non-parochial, or Other Denominations ($F(2, 30) = 2.48$, $p > .05$). Therefore, membership or categorization in a School Type was not a statistically significant factor that impacts school-level educational achievement. Statistically, classification by School Type accounted for 7.4% (η^2 = .074) of total variance in the School-Level Achievement estimates derived using Weighted-Indicator Scores Protocol, and 14.0% (η^2 = .140) of variance in the School-Level Achievement estimates derived via the Multilevel Achievement Estimation Protocol (Table 17).

Table 17

Differences between Model Estimates by School Type

	School Types	Sum of Squares	*df*	Mean Square	*F*	*p*	η^2
WISP	Between School Types	2.337	2	1.169	1.198	.316	.074
	Within School Types	29.266	30	0.976			
	Total	31.603	32				
		Sum of Squares	*df*	Mean Square	*F*	*p*	η^2
MAEP	Between School Types	3.957	2	1.978	2.448	.104	.140
	Within School Types	24.249	30	0.808			
	Total	28.206	32				

School-Level Achievement estimates derived from: Weighted-Indicator Scores Protocol (WISP), Multilevel Achievement Estimation Protocol (MAEP)

Eta squared (η^2) provides indication of total variance in the School-Level Achievement estimates associated with the School Type classification: Catholic, Non-parochial, or Other Denominations

The location of a school was revealed as having a statistically significant effect on the School-Level Achievement estimates, where a school was located mattered. There were statistically significant differences in the School-Level Achievement estimates derived using WISP ($F(2, 30) = 3.365$, $p < .05$). A school's location in Urban, Rural, or Remote Settings accounted for 18.3% (η^2 = .183) of the total variance in the estimate as determined in a one-way ANOVA (Table 18). For the School-Level Achievement estimates derived using the MAEP, the one-way ANOVA also revealed a statistically significant difference based on a school's location ($F(2, 30) = 4.494$, $p < .05$). Using the MAEP, it was determined that 23.1% ($\eta^2 = .231$) of the total variance was attributable to location: Urban, Rural, or Remote (Table 18).

A school's location mattered with statistical significance ($p < .05$) in terms of School-Level Achievement regardless of the protocol used. Subsequently, a Tukey honestly significant difference (Tukey HSD) post hoc test on the location was conducted to establish statistical significance by location category. The Tukey HSD revealed that the differences in School-Level Achievement were statistically significant between the Urban (higher) and Remote (lower) schools using both WISP and MAEP ($p < .05$) achievement estimates (Table 18).

For the achievement estimates from the Multilevel Achievement Estimation Protocol, there was also a statistically significant difference between the Rural and Remote locations ($p < .05$), but not for WISP ($p > .05$). Comparatively, for the schools located in urban and rural areas, there were no statistically significant differences with WISP estimates ($p > .05$) or with the MAEP estimates ($p > .05$) (Table 18).

Table 18

Differences between Model Estimates by School Location

		Sum of Squares	*df*	Mean Square	*F*	*p*	η^2
WISP	Between Location	5.790	2	2.895	3.365*	.048	.183
	Within Location	25.813	30	.860			
	Total	31.603	32				
MAEP	Between Location	6.503	2	3.251	4.494*	.020	.231
	Within Location	21.703	30	0.723			
	Total	28.206	32				

Tukey HSD Post Hoc Test

	(I) Location	(J) Location	(I-J) Mean Difference	*SE*	*p*
WISP	Urban	Rural	0.356	.379	.620
	Urban	Remote	1.125*	.440	.041
	Rural	Remote	0.769	.406	.158
MAEP	Urban	Rural	0.145	.347	.909
	Urban	Remote	1.113*	.403	.026
	Rural	Remote	0.968*	.372	.037

School-Level Achievement estimates derived from: Weighted-Indicator Scores Protocol (WISP), Multilevel Achievement Estimation Protocol (MAEP)

Eta squared (η^2) provides indication of total variance in the School-Level Achievement estimates associated with the location classifications: Urban, Rural, Remote

* Statistically significant

Therefore, the response to the second question follows that there were no differences in the School-Level Achievement estimates revealed between the three school types. However, statistically significant differences in the School-Level Achievement estimates existed in both the WISP and the MAEP estimates between the location groupings.

Question 3

How much variance in school-level academic outcomes was accounted for by selected Educator-Efficacy, Educators' Perceptions of Their School Environment, *and the Opportunity to Learn variables?*

The response to this question was broken down into three sets of analyses, one for each of the academic outcome variables addressed in this study: Primary School Examinations (Table 19), School-Based Averages (Table 20), and School-Rigor (Table 21). In each analysis (Stepwise), the independent variables (predictors) were Educator Efficacy, Opportunity to Learn, and Educators' Perceptions of Their School Environment. The "probability-of-F-to-enter ≤ .050, probability-of-F-to-remove ≥ .100" were the Stepwise criteria used. The Durbin-Watson test was used to detect autocorrelation, or similarities across series that can lead to underestimation of the standard error and false positive significance for predictors. The generally acceptable range for the Durbin-Watson test is between 1.5 and 2.5, with 2.0 indicating no autocorrelation, less than 2.0 as positive, and greater than 2.0 as negative (Glen et al., 2016). Multicollinearity, or the redundancy caused by predictor variables that are highly correlated or could be used to predict the other, was reported in terms of variance inflation factor (VIF). Uncorrelated predictors have VIF = 1.0; the generally acceptable range is between 1.0 and 5.0, while VIF > 5.0 indicates unacceptable highly correlated predictors (Glen, 2015).

Predicting Primary School Examinations

In the model with 2018 Primary School Examinations as the academic outcome variable, both Educator Efficacy and Educators' Perceptions of Their School

Environment did not meet the criteria (*p* > .05) and therefore were excluded. In the final model that met criteria for predicting Primary School Examinations, Opportunity to Learn was the only variable that predicted Primary School Examinations with statistical significance (R^2 = .199, *F* (1, 33) 8.175, *p* < .05) (Table 19). The model showed a slightly negative autocorrelation (DW= 2.3), which was within the acceptable range. No multicollinearity was detected (VIF = 1.00). Therefore, given the sample of schools in this study, Opportunity to Learn accounted for 19.9% of the total variance in School-Rigor with statistical significance. That Opportunity to Learn matters for school-level achievement was reflected in school-level averages for Primary School Examinations.

Table 19

PSE as the Academic Outcome Variable and EFF, OTL, and PRC as Predictors

Model Summary								
R	R^2	*Adj.* R^2	SE	*F*	*df1*	*df2*	*p*	*DW*
.446	.199	.174	.909	8.175	1	33	.007	2.313

Coefficients						
	Unstandardized	Standardized				
Model	B	*SE*	*Beta*	*t*	*p*	VIF
EFF^	.136			0.862	.395	.985
OTL	.446	.162	.446	2.859	.007	.100
PRC^	.018			0.095	.925	.683

Dependent variable: Primary School Examinations (PSE)

Predictor: Opportunity to Learn (OTL);

^Excluded (p>.10): Educator Efficacy (EFF), Educators' Perceptions of Their School Environment (PRC)

Method: Stepwise, n = 35

Predicting School-Based Averages

With School-Based Averages as the academic outcome variable, in the Stepwise regression analysis both Educator Efficacy and Opportunity to Learn variables did not meet the criterion (p > .05), thus excluded. In the model that was retained, autocorrelation and multicollinearity were detected, but within the acceptable ranges (DW = 1.56, VIF = 1.0). In this model to predict School-Based Averages, Educators' Perceptions of Their School Environment was the only statistically significant predictor variable (R^2 = .200, F(1, 34) 8.513, p < .05) (Table 20). Educators' Perceptions of Their School Environment accounted for 20.0% of the variance in School-Based Averages. Therefore, given the schools sampled in this study, Educators' Perceptions of Their School Environment matters for school-level educational achievement, as reflected in School-Based Averages.

Table 20

SBA as the Academic Outcome Variable and EFF, OTL, and PRC as Predictors

Model Summary

R	R^2	Adj. R^2	SE	F	$df1$	$df2$	p	Durbin-Watson
.447	.200	.177	.907	8.513	1	33	.006	1.558

Coefficients

Model	Unstandardized B	SE	Standardized Beta	t	p	VIF
EFF^	-.305			-2.021	.051	.946
OTL^	.147			0.774	.445	.661
PRC	.447	.153	.447	2.918	.006	1.00

Dependent variable: School-Based Averages (SBA)

Predictor: Educators' Perceptions of Their School Environment (PRC)

^Excluded: Educator Efficacy (EFF), Opportunity to Learn (OTL)

Method: *Stepwise*, n = 35

Predicting School-Rigor

In the third Stepwise regression analysis, in which School-Rigor was the academic outcome variable, neither Educator Efficacy nor Educators' Perceptions of Their School Environment met the input criteria. In the model that met the Stepwise criteria, autocorrelation was detected, but within the acceptable range (DW = 2.38). No multicollinearity was detected (VIF = 1.00). In this model, the Opportunity to Learn was the only variable of the three that predicted School-Rigor with statistical significance (R^2 = .173, $F(1, 33)$ 6.893, $p < .05$) (Table 21). In other words, given the schools sampled in this study, Opportunity to Learn accounted for 17.3% of the total variance in School-Rigor. Opportunity to Learn mattered for school-level educational achievement as reflected in School-Rigor.

Table 21

School-Rigor (RIG) as the Academic Outcome variable and EFF, OTL, and PRC as Predictors

Model Summary								
R	R^2	*Adj.* R^2	*SE*	*F*	*df1*	*df2*	*p*	Durbin-Watson
.416	.173	.148	.944	6.893	1	33	.013	2.368

Coefficients						
	Unstandardized		Standardized			
Model	B	*SE*	*Beta*	*t*	*p*	VIF
EFF^	.177			1.116	.273	1.015
OTL	.441	.168	.416	2.625	.013	.100
PRC^	-.036			-.183	.856	.683

Dependent variable: School-Rigor (RIG)

Predictor: Opportunity to Learn (OTL)

^Excluded: Educator Efficacy (EFF), Educators' Perceptions of Their School Environment (PRC)

Method: *Stepwise, n* = 35

Predicting Academic Outcomes

The variances in the three measures of academic outcome, Primary School Examinations, School-Based Averages, and School-Rigor, were not equally accounted for (explained) by the educator-related variables (Educator Efficacy, Opportunity to Learn, and Educators' Perceptions of Their School Environment). Opportunity to Learn was the best predictor of both Primary School Examinations and School-Rigor with statistical significance. The Primary School Examinations was an external measure reflecting the mean performance of eighth-grade students as a reflection of the school's achievement. School-Rigor was computed with consideration of both the Primary School Examinations and School-Based Averages (all grades). Therefore, for the schools sampled in this study, Opportunity to Learn explained a statistically significant portion of the variances in measures that included external assessments. However, the Educators' Perceptions of Their School Environment variable better explained variances in the School-Based Averages among the schools. The educators' self-reported ratings encompassing their internal capacity and likelihood to succeed in their job Educator Efficacy did not appear to significantly account for variances in academic outcome variables. School-level academic achievements, both externally and internally sourced, did not appear to be significantly impacted by Educator Efficacy for the schools represented in this study.

Question 4

Was there a difference between class averages for upper, middle, and lower divisions and school-level performance on the Primary School Examinations?

When determining if there were differences between class averages grouped by division (upper, middle, and lower), due to sample sizes, analyses were only possible using the overall data (district level). The analyses to answer this research question could include neither responses from preschool-only teachers nor those from educators who were not teachers. Sample composition also impacted the group sizes, which were unevenly distributed across school divisions: lower division (n = 59), middle division (n = 103), and upper division (n = 72). The harmonic mean was used in the Tukey HSD post hoc analyses; therefore, Type 1 errors (false positives) were possible.

As a whole, it was determined that there was a statistically significant difference for School-Based Averages estimates by class-level divisions ($F(2, 231) = 4.741$, $p < .05$, $\eta^2 = .039$) (Table 22). Class-level divisions accounted for 3.9% of the total variance in School-Based Averages. In the multigroup comparisons (post hoc test), the only statistically significant difference for class averages by divisions was revealed between lower and the upper divisions ($M_{diff} = 0.565$, $SE = 0.186$, $p < .05$) (Table 22). For the Primary School Examinations, it was determined that there were no statistically significant differences in estimates by class-level divisions ($F(2, 229) = .020$, $p > .05$, $\eta^2 < .001$) (Table 22). Since division was not a significant predictor of Primary School Examinations, a school-level variable, no further post hoc tests were conducted.

Table 22

Differences between School-Based Averages by Class Level/Divisions

ANOVA: Model Estimates by Division							
		Sum of Squares	*df*	Mean Square	*F*	*p*	η^2
SBA	Between Division	10.662	2	5.331	4.741	.010	.039
	Within Division	259.738	231	1.124			
	Total	270.400	233				
PSE	Between Division	0.040	2	0.020	0.020	.981	0
	Within Division	234.711	229	1.025			
	Total	234.751	231				

Tukey HSD Post Hoc Test					
	(I) Division	(J) Division	(I-J) Mean Difference	*SE*	*P*
SBA	Lower	Middle	0.386	.173	.069
	Lower	Upper	0.565*	.186	.008
	Middle	Upper	0.179	.163	.517

Dependent variable: School-Based Averages (SBA)

Independent variable: Division (Lower, Middle, Upper)

* Statistically significant at the .05 level.

Therefore, the fourth question about differences in School-Based Averages and Primary School Examinations by division revealed that overall there was a significant difference in School-Based Averages by division, but the difference was only statistically significant between upper and lower divisions. Overall, the class division was not a significant predictor of Primary School Examinations.

Question 5

What was the relationship between selected school characteristic variables (School Size, Location, and Infrastructure) and variables for Educator-Efficacy and Educators' Perceptions of Their School Environment?

The relationships between school characteristics and educator-related variables were assessed and described in terms of Pearson correlation coefficients and *p*-values (Table 23). The analyses included two measures for the School Size: student population and the number of educators in schools. In addition to School Size, there were two other school characteristic variables, Location and Infrastructure. The educator-related variables of particular interest were Educator Efficacy and Educators' Perceptions of Their School Environment of the school environment.

At the school level, the correlational analyses (Table 23) indicated that there were negative associations between school-size variables (student population, number of educators) and educator-related variables (Educator Efficacy and Educators' Perceptions of Their School Environment). In school-level comparisons, the relationships were not statistically significant ($p > .05$) between School Size and Educator Efficacy or between School Size and Educators' Perceptions of Their School Environment (Table 23).

In other school-level correlational analyses, the relationship between a school's location and educator-related variables presented negative coefficients. The negative coefficient for the relationship between the school's location and Educator Efficacy was not statistically significant ($p > .05$). However, the negative relationship between Location and Educators' Perceptions of Their School Environment was statistically significant (r (35) = -.492, $p < .001$) (Table 23). The Location grouping codes were, Urban = 1, Rural =

2, and Remote = 3; therefore, Educators' Perceptions of Their School Environment were less favorable for schools located farther away from the urban areas.

At the school level, Pearson correlation analyses indicated that there were statistically significant positive associations between the physical attributes of the school's infrastructure and the educator-related variables. The relationship was statistically significant between physical attributes of the school's infrastructure and Educator Efficacy ($r(35) = .486$, $p < .05$). Similarly, the Pearson correlation analyses showed that the relationship between the physical attributes of the school's infrastructure and the Educators' Perceptions of Their School Environment ($r(35) = .398$, $p < .05$) was statistically significant.

Table 23

Correlations between Select School-Characteristic Variables and Educator-Related Variables

Data Level				Educator Variables		School Characteristic Variables			
		M	*SD*	EFF	PRC	Size-E	Size-S	LOCN	PHYS
School	EFF	386.97	26.35	1	.519**	-.189	-.243	-.149	.486**
	PRC	377.70	40.57		1	-.218	-.205	-.492**	.398**
	Size-E	12.44	7.45			1	.951**	.183	-.189
	Size-S	259.81	190.16				1	.229	-.200
	LOCN	112.50	30.18					1	-.295*
	PHYS	61.70	12.74						1
Overall (District)	EFF	384.21	54.68	1	.510**	-.105*	-.115*	-.082	.225**
	PRC	375.72	65.36		1	-.091	-.093	-.379**	.303**
	Size-E	16.49	9.32			1	.969**	.171**	-.099*
	Size-S	370.21	245.25				1	.207**	-.101*
	LOCN	108.25	30.04					1	-.282**
	PHYS	60.12	22.09						1

** Correlation is significant at the 0.01 level (1-tailed).

* Correlation is significant at the 0.05 level (1-tailed).

School-Level: Listwise $n = 36$

Overall: Listwise $n = 294$

Overall, at the district level, Pearson correlation coefficients did not reveal any statistically significant relationships between School Size and Educators' Perceptions of Their School Environment. However, there were statistically significant negative associations at the .05 significance level between Student Population (Size-S) and Educator Efficacy ($r(294) = -.115$, $p < .05$). Similarly, the relationship between number of educators (Size-E) and Educator Efficacy ($r(294) = -.105$, $p < .05$) was statistically significant (Table 23). Correlational analysis also revealed positive statistically significant associations between physical attributes of the school's infrastructure and Educator Efficacy ($r(294) = .225$, $p < .001$) and between physical attributes of the school's infrastructure and Educators' Perceptions of Their School Environment ($r(294) = .303$, $p < .001$). Furthermore, the relationship between Location and Educators' Perceptions of Their School Environment ($r(294) = -.379$, $p < .001$) was statistically significant, but the relationship between the school's Location and Educator Efficacy was not significant.

In response to the fifth question, there were statistically significant relationships between select school characteristic variables and educator-related variables at the school level and the overall district level. There were positive statistically significant associations for the physical attributes of schools and both educator-related variables in every comparison. Where a school was located (Location variable) also revealed statistically significant associations, at both levels, with the Educators' Perceptions of Their School Environment. Correlations between the school-size variables and educator-related variables revealed only one significant association with Educator Efficacy at the district level and none at all with Educators' Perceptions of Their School Environment.

Overall at the district level, for the schools included in this particular study sample, School Size appeared to have a proportional impact on the how educator's perceived their schools, but that relationship changed when considered at the school level. What seemed to have had a greater impact was the condition or physical attributes of the schools.

Chapter 5

Discussion

Overview

This chapter discusses the five primary research questions associated with this study, the data that was used, and some considerations for the Belize education system. The first question in this study asked if there were differences between the School-Level Achievement estimates, as derived from the Base-Model using the Weighted-Indicator Scores Protocol and the Multilevel Achievement Estimation Protocol. The remaining questions sought to examine specific relationships and interactions among variables and estimates. In this study, the concept and estimation of School-Level Achievement focused on School Context, Educator Efficacy, Opportunity to Learn, Educators' Perceptions of Their School Environment, School-Based Averages, School-Rigor, and school averages on the 2018 Primary School Examinations. The *Belize Educator Survey* served as the primary source of data collected directly from educators in one school district of Belize.

Research Questions

Question 1

Is there a difference in the overall School-Level Achievement estimates using a classic summative approach with the Weighted-Indicator Scores Protocol (WISP) versus a latent approach with a Multilevel Achievement Estimation Protocol (MAEP)?

The fundamental differences in the methodology between the Weighted-Indicator Scores Protocol and the Multilevel Achievement Estimation Protocol estimation variables were the approach and the underpinnings of the final computation of the five overarching estimation variable scores and the School-Level Achievement estimates. In the approach for the Weighted-Indicator Scores Protocol, more steps were involved in the process to determine the weights as compared to the Multilevel Achievement Estimation Protocol approach.

Further model comparisons used standardized estimates (*z*-scores), general descriptive statistics, rank-order comparisons, *t*-tests, correlations, and multiple regression analyses to tease out differences between the models. The differences found in the descriptive statistics for the standardized estimates were minimal between the standardized model estimates and Primary School Examinations. Minimal differences also surfaced between the Weighted-Indicator Scores Protocol and Multilevel Achievement Estimation Protocol standardized estimates with consideration at the district level and the school examples. A *t*-test of the standardized estimates also determined that the differences between the Weighted-Indicator Scores Protocol and Multilevel Achievement Estimation Protocol estimates were not statistically different.

The next level of differentiation was to determine which model estimates were better able to predict Primary School Examinations, the known measure of academic achievement for eighth-grade students. The determination was consistent with indications from previous analyses; the differences between the two sets of estimates were not statistically significant enough to merit one over the other. From an operational standpoint, neither method was superior. There were more steps involved in Weighted-Indicator Scores Protocol than the Multilevel Achievement Estimation Protocol. However, it was more challenging to fit the model for the Multilevel Achievement Estimation Protocol. Both methodologies represent the voices of educators, one of the most distinct, meaningful, and appealing factors of the effort to estimate School-Level Achievement in Belize.

Question 2

Were there differences in the School-Level Achievement estimates by selected school characteristics such as School Type (Catholic, Non-parochial, or Other Denominations) or Location (Urban, Rural, or Remote)?

ANOVA (analysis of variance) was used to determine if a school's address (Location) or the management (School Type) impacted achievement. A school's designated location was either Urban, Rural, or Remote. Location classifications were based on the distance from an urban center, road conditions, access to infrastructural utilities, and availability of public transportation to the school. There were no statistically significant differences in the school-level estimates by estimation method—the Weighted-Indicator Scores Protocol or the Multilevel Achievement Estimation Protocol. However, analyses revealed that location made a statistically significant difference with the achievement estimates. This finding indicated that where students live and attend school make a difference, and the difference was more favorable for those who went to school in the urban and rural areas (as opposed to remote) in this school district.

The Opdenakker and Van Damme (2007) study indicated that School Type (Catholic versus non-Catholic) mattered in their study conducted in Flanders, Belgium. However, the research conditions between Western Europe and Belize were very different. The point was to see if School Type was a common underlying factor that impacted achievement. In this study School Type was predetermined by school management; therefore, schools were categorized as Catholic, Non Parochial (government and community), or Other Denominations. Unlike that study in Flanders, Belgium, School Type did not appear to have statistically a significant effect on the School-Level Achievement estimates for this particular sample in Belize.

Question 3

How much variance in school-level academic outcomes was accounted for by selected Educator Efficacy, Educators' Perceptions of Their School Environment, and the Opportunity to Learn variables?

Recap of Academic Outcomes

The general measures of school-level academic outcomes were School-Based Averages, Primary School Examinations averages, and School-Rigor. Primary School Examinations are standardized tests prepared by the Belize Ministry of Education for eighth-grade students in English, Mathematics, Science, and Social Studies. Combined, the scores from these tests are used as a measure of academic achievement. The tests were written for the country rather than tailored to each specific school. Therefore, the Primary School Examinations could cover academic materials to which some students were never exposed. The educators in the primary schools do not have control over development or administration of the item bank. On the contrary, School-Based Averages in this study were cumulative and school specific, based on classroom assessments, and the individual educators were in control of their student assessments (tests, quizzes, homework, projects, participation, etc.). With the sample in this study, there were differences in how much variance was accounted for with the Primary School Examinations averages, School-Based Averages, and School-Rigor as measures of academic outcomes with the three educator-informed variables (Educator Efficacy, Opportunity to Learn, and School-Based Averages). In a nutshell, Primary School Examinations are national standardized tests for one grade while the School-Based Averages are derived from educator-specific assessments in each grade.

Estimation Variables and Academic Outcomes

In this study, one finding was that Opportunity to Learn was the only estimation variable that appeared to account for variance with statistical significance when the academic outcome variable was either Primary School Examinations or School-Rigor. Educator Efficacy levels had a marginal effect on the School-Based Averages, but the Perceptions of their School Environment variable accounted for variance with statistical significance when the academic outcome variable was School-Based Averages.

Opportunity to Learn involved some factors that were beyond the control of educators, such as language, socioeconomics, access to learning materials, factors in the family and community, student-teacher ratios, and curricula. Other factors, such as time investments, educator preparedness, instruction time and content, attention and student-teacher interaction, and class discipline, were within the educators' realm of influence to increase students' opportunities to engage and learn. The significant result for Opportunity to Learn as a predictor of performance on Primary School Examinations, but not as a significant predictor of School-Based-Averages, is worthy of further investigation, especially to determine how much the difference had to do with the fact that one outcome was based on student performance in a single grade and the other was school-wide. More narrowly focused studies of how specific Opportunity to Learn facets interact are warranted and encouraged for the Belize Ministry of Education and local school managers. A clear understanding of the intricacies of Opportunity to Learn variables within the Belizean context could guide intervention measures to reduce potential disparities that stem from the issues or conditions that might be beyond the control of educators.

The Educators' Perceptions of Their School Environment apparently mattered significantly for academic outcome that was based on the assessments that were within the educators' control. That educators' perceptions could make a difference on the academic achievement for this study sample also aligned with findings from O'Malley et al. (2014). Their study showed that efficacy (students' and educators') and perceptions (about the school) impact how a school functions and achieves (O'Malley et al., 2014). Both Opportunity to Learn and Educators' Perceptions of Their School Environment were significant predictors of academic outcomes (standardized tests and school-based measures); therefore, those two variables seem to have the potential to impact students' success.

The focus on the educators' voice (Educator Efficacy and Educators' Perceptions of Their School Environment), Opportunity to Learn, School Context, and the use of School-Based Averages to estimate School-Level Achievement sets this study apart from other studies. In this study, because School-Based Averages were used, all educators and students (academics) in a school contributed to the estimation of School-Level Achievement and the story of their school. Weisberg et al. (2009) posited that "in a knowledge-based economy that makes education more important than ever, teachers matter more than ever" (p. 2).

Question 4

Was there a difference between class averages for upper, middle, and lower divisions and school-level performance on the Primary School Examinations?

In this study, there were unbalanced sample sizes across the divisions. Given the other data limitations previously noted, this question was considered at the district level only. The averages for the Primary School Examinations characterized achievement for eighth-grade students. When the School-Based Average variable was aggregated by division (upper, middle, and lower) in the district, the differences were not profound and did not appear meaningful. Statistically significant differences appeared between the upper and lower divisions in schools only, which does not seem meaningful. The differences were most likely attributable to naturally occurring developmental differences in maturity, development, basic educational experiences, and grade-level expectations.

As discussion prompts, the differences in School-Based Averages by division or which division best predicts Primary School Examination scores might be engaging among educators. However, given the sample size and the data in this study, there were no meaningful differences by divisional levels in school to predict Primary School Examinations, and further inferences would not be advisable.

Question 5

What was the relationship between selected school characteristic variables (, location, and infrastructure) and variables for Educator Efficacy and Educators' Perceptions of Their school environment?

The two School Size variables in this study were references to population counts not infrastructure—the number of students and the number of educators. On the surface, the strong relationship would not seem very meaningful, considering that both variables essentially described an aspect of population count. However, within the context of the Belize education system, this strong correlation could be interpreted as validation or confirmation of compliance. That is, on average, the school district sampled complied with the allocation proportion formula in the school year this study was conducted. The formula was set by the Belize Ministry of Education, and it is intended to regulate the distribution of educators assigned to schools based on the student population, with no special favor to any particular type of school. The formula also specifies to district managers the recommended number of administrators in a school and whether a principal leads a school with dual roles (teacher and administrator).

The relationships between Educator Efficacy and Educators' Perceptions of Their School Environment affirmed the interconnectedness of how educators see their school environment and how educators feel about their capacity to achieve success in their profession. The issue of Educators' Perceptions of Their School Environment is a matter related to the structure of the Belize education system and who holds the ultimate responsibility and accountability for the upkeep of learning spaces. The salaries for most educators come from public funds. Although wages are the most substantial part of the education budget—more than 20% of the national public expenditure (Arcia, 2016)—there are other substantial costs associated with operating a school. In the church-state partnership for education, the nongovernment school management entities are technically responsible for costs related to the upkeep of the school.

In this study, Educators' Perceptions of Their School Environment was conceptualized as a complex estimation variable that represented more than the obvious physical image of the school. The educators' perceptions correlated significantly with both location (negatively) and physical condition (positively) of the school, which could possibly be explained as a function of correlated attitudes or biases. The significant relationships, both negative and positive, affirmed the complex nature of educators' perceptions. The estimation variable captured more than just the physical condition in the collective voice of educators' perceptions. Conversely, it also illustrated how difficult it could be to compartmentalize variables. In general, the results pertaining to the educators' perceptions, their voice, contributed much as a measured aspect in the estimation of School-Level Achievement in this study.

Core elements of the Opportunity to Learn construct are instructional content, instructional quality, and time on learning (Banicky, 2000; Elliott & Bartlett, 2016). Student-teacher ratios are used to estimate the average time that educators potentially have for individualized attention with students. The number of students and teachers, or the ratios thereof, represent the number of classmates with whom students might have the opportunity to learn, play, and build relationships. By that same logic, increased student-teacher ratios would technically reduce opportunities for one-to-one interactions with the educator. The findings in this study suggested that among the educators in this school district, Educator Efficacy levels are likely to be higher in smaller (fewer students) schools than in larger (more students) schools. Given that Educator Efficacy is relevant for School-Level Achievement, such findings would be relevant for policymakers and school managers who advocate for smaller class sizes or lower student-teacher ratios.

General Discussion about the Models

The unweighted standardized estimation variable scores resulting from the *Belize Educator Survey* became, in effect, the Base-Model. The Base-Model served as the common base (initial data source) for both estimation protocols. The types of analyses used in the Weighted-Indicator Scores Protocol were inarguably classic, having been established before the middle of the 20th century. Basically, this approach estimated the amount of variance explained for each *estimation* variable and applied weights to the Base-Model. Conversely, the Multilevel Achievement Estimation Protocol accounted for the nested nature of the sample with a more contemporary iteration of multilevel structural equation modeling approach published in the 21st century. Though the approaches in the estimation protocols were different, the overall comparisons did not show differences that were statistically significant between the subsequent estimates for School-Level Achievement. Both the Weighted-Indicator Scores Protocol and the Multilevel Achievement Estimation Protocol involve analyses appropriate for small such of the population demographics in Belize. Therefore, either or both protocols could be used to estimate School-Level Achievement within the education system in Belize.

The Survey

The *Belize Educator Survey* was administered shortly before the two-week Easter break, which in retrospect, was less than optimal timing. The educators were asked to provide information about the previous academic year, and at that point, some educators were working in different schools. Others no longer had access to their grade books for the past year. Nevertheless, the first round of data collection exceeded the minimum threshold of participants for continuing this study without the need for subsequent rounds. However, successfully implementing actual full-coverage data collection where every primary school educator participates could be a realistic expectation with systematic implementation by the Belize Ministry of Education. Systemic full-coverage data collection would have increased the richness and the relevance of profiles elaborated for the schools and the district.

Conceivably, using the *Belize Educator Survey* as a primary data collection instrument to estimate School-Level Achievement may be a feasible alternative to large-scale test-based accountability in terms of time investment, financial cost, environmental impact, and benefit (data). First, it would take less time and resources to improve the *Belize Educator Survey* compared to what it would take to create multiple new standardized student tests. Second, because there are fewer educators than students, it would cost the government less to physically produce and process the *Belize Educator Survey* than to develop a new, large-scale standardized test for students (K–8). The third and often ignored aspect is the impact on the environment. The use of the *Belize Educator Survey* with all educators would impose a smaller carbon footprint (require less paper in a nation with limited recycling facilities) than standardized tests for all primary school students (K–8).

Another advantage of using the *Belize Educator Survey* instead of large-scale standardized testing is that the survey collects from educators, data that would not be appropriate or possible to obtain from students. In addition to the class averages of grades students earned in four subjects, the survey in this study captured information from educators about efficacy, perceptions, the school, and opportunities. The class averages were not standardized, but they represent how groups of students performed, given their circumstances over some time. On the other hand, a standardized test would have continued an established mode for measuring performance (single point in time) with the same metrics and protocols across schools and districts. Standardized testing would likely produce student-earned scores that would be comparable and recognized because they would be from a traditional source.

Educator Input

The overarching approach for estimating School-Level Achievement in this study was recognizing educators as the experts in their part of the education system. As noted in Chapter 2, the processes for how educational achievement was measured and how outcomes were expressed indirectly conveyed a sense of what goals and values were important for those who endeavored to quantify or qualify achievement (Cole, 1990). In the development of the Belize Educator Survey, sensitivity to the cultural nuances influenced the content, process, and revisions. The survey could become the standard measure applied to all educators (teachers and principals) in the country at the end of each academic year.

As mentioned in the literature review (Chapter 2), multiple studies have indicated that educators are the experts in their respective schools and the educators influence the narrative for success in the school individually and collectively (Caprara et al., 2006; O’Malley et al., 2014). Furthermore, educators’ beliefs and perspectives also play a significant role in both job satisfaction and academic achievement in other settings (Caprara et al., 2006; Heyneman & Loxley, 1983). As such, in this study, it was pivotal to tap into the educators’ collective voice to create the estimation variables and to show educators that their contribution matters for their students as well as for the story told of their school (School-Level Achievement).

By having educators as a primary source of data, the optics of credibility and educator-buy-in for the estimation process and outcomes would be more likely (Deal & Peterson, 2009). The School-Level Achievement estimates are likely to draw attention to these five issues that theoretically influence achievement: (a) School Context, or characteristics, (b) Educator Efficacy, (c) Opportunity to Learn, (d) Educators' Perceptions of their School's Environment, and (e) School-Based Averages of student performance based on teacher-prepared assessments over time (Agasisti et al., 2018; Bandura, 2014; Banicky, 2000; Caprara et al., 2006; Elliott & Bartlett, 2016; Fullan et al., 2015; Heyneman & Loxley, 1983; Masinoa & Niño-Zarazúa, 2016). Attention to these matters might initiate conversations among educators, administrators, and policymakers, the kinds of dialogue that can influence changes that improve the processes and experiences associated with education in Belize.

Academics and the Concept of School-Level Achievement

In this study, School-Level Achievement was a multifaceted concept. Therefore, there were five Base-Model factors of School-Level Achievement: School Context, Educator Efficacy, Opportunity to Learn, Educators' Perceptions of Their School Environment, and School-Based-Averages. The school-level variable School-Rigor was contingent upon the performance on the Primary School Examinations (standardized examinations for eighth grade only) and the educator-generated School-Based Averages (not standardized and includes all grades).

This study's approach and methods drew from the premise that grade-point averages, or School-Based Averages for this study, are cumulative expressions of academic performance via multiple subject areas, types of assessments, and a history that contributed to the academic experience and performance. Conceptually, School-Based Averages represent to the school what grade-point averages are for students. This study sought to include educators' assessment of students' academic performance by using the overall class averages reported for all grade levels (K–8) to compute School-Based Averages. Systemic differences of assessments in schools (K–8) and between schools were acknowledged. The schools in this study used different criteria and standards for educational assessment, which in most instances diminished comparability. However, School-Based Averages as an outcome estimation variable for academic performance, as used in this study, paralleled grade-point averages for students. Both measures represent how the unit performed on academics in the context of the system or community in which it existed. In this School-Level Achievement model, academic performance was only one of five estimation variables for School-Level Achievement, which signified that academics was necessary but not sufficient to demonstrate School-Level Achievement.

Considerations for Estimation

The Base-Model variables were evaluated in terms of representativeness, internal consistency, and integrity of the data used in this study. The psychometric properties of the data appeared satisfactory, which in some ways endorsed the quality of the data-collection instrument and the sample. The associated logistics and demographics suggested that the representativeness of the sample was appropriate and ample for this study.

The Base-Model served as the foundation for the two models tested in this study. The data distribution and variability of the Base-Model were found to be reasonably normal, and the correlations between estimation variables supported the theoretical structures for the variables. As a methodological caveat, it should be noted that when correlations both come from the same study, there is generally an inherent correlation because the data were contributed by the same people. Nevertheless, that the data collected were of reasonable quality increases confidence for inferences, and the overall results were positive for the Base-Model. Therefore, to the policymaker or the administrator, the psychometric results for the Base-Model—scientific evidence—can affirm outcomes and inform decisions.

For the Weighted-Indicator Scores Protocol model, it was important to assess the dimensionality and amount of variance explained by the factors because those statistics informed the weights later applied to the variables. Confirmatory factor analyses determined that all the variables had a dominant first factor that accounted for more than 20% of the variance, which is the minimum threshold noted in Reckase (1979) and in Zopluoglu and Davenport (2017). Notably, the Opportunity to Learn variable, which underwent the most revision after the field test (pilot), improved considerably. The Opportunity to Learn concept was described in the literature as complex and multifaceted (Banicky, 2000; Schuh-Moore et al., 2012). Opportunity to Learn factors were not limited to the students' socioeconomic status or the type of school. Opportunity to Learn factors were connected to the community, the family, the instruction content and process, the educator's qualification, the students' motivation, and many other factors or facets that contribute to academic outcomes (Banicky, 2000; Schuh-Moore et al., 2012). The

Opportunity to Learn variable revealed the most (seven) factors or dimensions extracted, compared to all estimation variables in this study. The multidimensional extraction for the Opportunity to Learn variable was neither unexpected nor alarming. In this study, 23 items composed the Opportunity to Learn variable. In the revision process, the items were deliberately focused into one of several specific aspects of the overarching theme, Opportunity to Learn. Collectively as an estimation variable, the various aspects were meaningful and intended to be holistic.

The literature reviewed in Chapter 2 presented Opportunity to Learn as a multidimensional or multifaceted concept (Banicky, 2000; Heafner & Fitchett, 2015; Schuh-Moore et al., 2012). There are multiple factors associated with the construct for Opportunity to Learn. In this study, the 23 items that informed the Opportunity to Learn variable included items that addressed facets described in other studies, such as designated instruction time, access, resources, behavior, the curriculum, student-teacher ratios, whether students have necessary materials, language, and attendance (Banicky, 2000; Elliott & Bartlett, 2016; Fullan et al., 2015; Grantham-McGregor et al., 2007; Heafner & Fitchett, 2015; Masinoa & Niño-Zarazúa, 2016; Schuh-Moore et al., 2012; Segretin et al., 2016). In retrospect, rather than compiling so many variables into a single estimation variable, it might have been more prudent to split the variable into several narrowly focused constructs that theoretically contribute meaningfully to the concept. Therefore, the results of this study affirm the multifocal conceptualization of Opportunity to Learn, and that alone was a positively meaningful finding (Banicky, 2000; Elliott & Bartlett, 2016). Evaluating and enhancing opportunities for students to learn is not a simple task, and subsequently, solutions or strategies to address deficits require multifocal approaches. As described in the literature review, the overarching or umbrella concept of Opportunity to Learn involves an array of issues related to access, inputs, and processes that impact achievement (Banicky, 2000; Elliott & Bartlett, 2016).

For future use, there is still room for improvement with the internal consistency statistics and dimensionality for the Opportunity to Learn variable evidenced in the overall coefficient alpha (.734) achieved in this study (Table 4). Through the educators' responses, the estimation variable conceivably captured a general impression of access, resources, socioeconomic status, behavior, and curricula. Strengthening this particular estimation variable would require more items and compartmentalizing the Opportunity to Learn concept into multiple estimation variables. Having a more specific focus might help to differentiate relevant strengths and weaknesses across schools or school districts.

In tandem with the Heyneman-Loxley effect (Heyneman & Loxley, 1983; Zumbach, 2010), the total variances explained for Opportunity to Learn and School Context showed that both variables matter to school-level educational outcomes. One of the reasons for estimating achievement was to initiate positive changes and meaningful conversations about what could be impacting the lower scores observed in the Primary School Examinations across the district. The estimation variables used in this study could be that starting point. As previously indicated, the Base-Model variables for this study provided the foundation for both Weighted-Indicator Scores Protocol estimation and the Multilevel Achievement Estimation Protocol. The Base-Model variables were deemed representative, reliable, and structurally sound and had the statistical integrity for the purposes of this study.

Conceptually, the Weighted-Indicator Scores Protocol achievement estimates were comparable to final grades of a course. Scores, proportional (weighted) to perceived value in multiple assignments, tests, and projects (academic performance indicators), were combined and reported as a single grade or average. Simply put, the Weighted-Indicator Scores Protocol, combined five relevant components as estimation variables that were weighted according to variance (variability explained/accounted for). The Weighted-Indicator Scores Protocol can be simplified in terms of the overarching concepts associated with weighting grades according to levels of difficulty, content relevance, time investment, or creativity for a course or a grade level. The technical aspects in the Weighted-Indicator Scores Protocol, although based on classic statistical analyses, may or may not be as easily understood by some educators. Similarly, the Multilevel Achievement Estimation Protocol model input and estimations might appear obscure or overly technical to a primary school teacher, especially one who has never taken a statistics class. Conceptually, to demystify the Multilevel Achievement Estimation Protocol for the educators, the overarching parallel for the multilevel approach can be an organizational chart and consideration of the connections elements and/or contributions from different parts that have an impact on whether the organization succeed or fail. Some qualities or driving forces in the organization are not tangible—the latent factor in the Multilevel Achievement Estimation Protocol.

Limitations

One of the limitations of this study was the missing data (class averages) in the educators' responses in the *Belize Educator Survey*. The work-around to complete the data set and test the estimation models incorporated stochastic imputation combined with random weighted probability. With no intent to link the estimates to the actual participating schools and a focus on comparing estimation methods, stochastic imputation combined with random weighted probability was most likely the most efficient and effective study-specific option for addressing missing data (Burns, 2019; Scholtus et al., 2014; Seaman et al., 2012).

The stochastic imputation combined with the random weighted probability resulted in data that showed variability that mimicked the descriptive properties of the data for each school. However, the work-around method used to address missing data appeared appropriate within the scope of this study but would not be encouraged nor recommended for operational implementation. Instead of a work-around for missing data, the more prudent course of action would be to initiate the round-two data-collection protocols and follow-up directly with educators or schools to collect a complete or highly representative sample of the academic data. Inevitably, decisions for appropriately addressing missing data would then depend on the magnitude of "missingness" after follow-up protocols are exhausted.

An inherent limitation for studies conducted in Belize is population size. Belize is not heavily populated (fewer than 400,000 people), but Belize is culturally and linguistically diverse. Nevertheless, the results of the psychometric analyses of data collected in this study were more than adequate. However, those results were specific to a sample from one district in education system and should not be generalized beyond that.

Another significant limitation with this study was the narrowed scope for generalization given the relatively small sample size. The findings for this study should not be extrapolated or extended beyond Belize. Even within the country, it is not advisable at this point to generalize beyond the district sampled.

Summary

This study examined two methods for estimating School-Level Achievement in Belize. In tandem with the overarching concept of school-level educational achievement, this study considered the voice of educators (Educator Efficacy and Educators' Perceptions of Their School Environment), School Context, Opportunity to Learn, and School-Based Averages. This study contributed a culturally adapted data-collection instrument, two viable methods to estimate School-Level Achievement, and it introduced an approach for framing School-Level Achievement. The approach used in this study differed from others because it valued the input of those who operationalized the education process—the primary school educators. School-Level Achievement estimates were a function of five variables, not solely academics.

Standardized tests are typically the driving force of accountability and achievement in test-based systems. The intent of this study was not to debate the importance or the value of using standardized tests. In the Belize education system, which has few applications of large-scale standardized tests, consideration (or innovation) of other options was imperative. Hence, this study used a concept of School-Level Achievement that included academic assessment as one of five contributing factors in the overall estimation, but not as the only or the primary source of information. The Primary School Examination was the singular standardized assessment available as a reference point. Therefore, the primary measure of academic achievement for this study was the cumulative measure of student performance from School-Based Averages collected as class-level grades from the educators responsible for the grades. This study aimed to capture the voices of educators, particularly in their evaluation of their school environment, their efficacy, and the education process.

The estimates from both methodologies compared in this study, the Weighted-Indicator Scores Protocol and the Multilevel Achievement Estimation Protocol, indicated that School Type was not a statistically significant as a factor in School-Level Achievement estimates. That suggested that in the church-state public school system that existed in the school district sampled, no particular advantage came from attending a school managed by the government, a community, or a church. (Further study is recommended.)

This study also affirmed the relevance of Educator Efficacy and Opportunity to Learn to academic performance. The Educator Efficacy estimation variable predicts student performance on internal assessments, while Opportunity to Learn predicts performance on Primary School Examinations. School administrators and policymakers should identify and try to improve or address issues that may detract from educators' efficacy and students' opportunities to learn. It is important for educators to realize that as agents of change, their perceptions of their school and their individual efficacy may be key factors toward students' accomplishments. Therefore, educators' voices matter. Educators can assert positive changes in themselves to impact school-level success.

In the absence of formal or established means of estimating School-Level Achievement, as a result of this study, there are now two viable options: Weighted-Indicator Scores Protocol and Multilevel Achievement Estimation Protocol. What could those particular options mean for the Belize education system? Either Weighted-Indicator

Scores Protocol or Multilevel Achievement Estimation Protocol can be used in conjunction with the *Belize Educator Survey* and most likely produce meaningful estimates to inform school-based strategies and policies. Ultimately for Belize, the choice between Weighted-Indicator Scores Protocol and Multilevel Achievement Estimation Protocol may not be a matter of statistics but rather one of trust (face validity) and user preference or access to the software options. Both methodologies were considered appropriate for application in the Belize education system. In either case, the School-Level Achievement estimates were inclusive of five estimation variables and, therefore, the educators' voice, the school setting, and students' grades. To reiterate a previous statement, the most fundamental difference between the two methods was that the Weighted-Indicator Scores Protocol approach considered how much variance was explained by variables to formulate weights using analyses that are more common and time-tested in the field (classic). The Multilevel Achievement Estimation Protocol approach considered measurement error or accuracy and used an estimation process with very complex computing that might seem obscure to those who are unfamiliar with quantitative methodology.

This study added to the descriptions of relationships that are worthy of future studies and efforts needed to develop sustainable strategies and to address specific issues. The voice of educators in measures may have introduced some bias associated with self-report, but the level of meaningfulness that was gained in authenticity, relevance, and school-specific experiences was invaluable. In this study, there was a strong relationship between educators' efficacy and their perceptions of the school environment. Educators' Perceptions of Their School Environment decreased with increased School Size and increased distance from the urban areas. However, the Educator Efficacy variable did not correlate significantly with the location. That finding suggested that a school’s address was not likely to have a direct effect on how educators feel about being able to do their job. How educators perceived their environment was more likely to affect their efficacy.

As a result of this study, School-Level Achievement estimation variables and the overall estimates might assist others in identifying strengths and weaknesses within schools. These methodologies rely on input from experts in the schools, the educators. School management, principals, educators, and the Belize Ministry of Education can use components of the School-Level Achievement estimates to inform the development of specific measures or strategies to improve the education process. As a data- collection tool, the *Belize Educator Survey* captures reliable and valid information for the intended purpose (to quantify the educators’ perspectives for five estimation variables in School-Level Achievement) and the intended population (the residents of Belize). This survey instrument was developed specifically for use in the Belize education system. The survey yielded data used to compute the variables that comprised various estimation variables, and the remaining data comes from existing records shared by the Belize Ministry of Education.

Beyond the scope of this study, there are two key next steps to complement the school-level educational achievement estimation process. The first would be to finalize a formal format to report School-Level Achievement or school profile. A suggestion is to solicit input from the education commission and educators (small groups, district level, and management levels) to explore platforms and formats to disseminate information according to stakeholders’ needs. The second would be to identify and/or develop additional measures for academic performance. The instrument should align with grade-level expectations in the Belize education system. The instrument's design should also ensure that every educator has the opportunity to contribute and that all primary schools have the chance to be included in estimations and be appropriate for the smaller and nontraditional schools.

The purpose of this study was to develop a method for estimating School-Level Achievement, and ultimately two methods or protocols resulted from this study. In addition to estimates, individual scores from the estimation variables can also be used to create school-level characterizations or profiles. This study also achieved the goal to include the collective voice or perspective of educators in terms of Educator Efficacy and Educators' Perceptions of Their School Environment. Educators are among the local experts of their school; thus, their perspectives should contribute considerably to the overall impression sketched (estimated) for their school. Nevertheless, within the concept of School-Level Achievement in this study, the educators' voices (Educator Efficacy and Educators' Perceptions of Their School Environment) were considered to be "necessary but not sufficient" to constitute achievement. The School-Level Achievement estimates and the foreseen school profiles would include all five estimation variables. The estimates and school profiles can serve the school district and schools as tools to identify schools that could be partnered for mutually beneficial activities or to pinpoint case-specific opportunities for growth.

In conclusion, this study presented a conceptual model for School-Level Achievement and a data-collection instrument. The model recognized five estimation variables for School-Level Achievement: School Context, Educator Efficacy, Opportunity to Learn, Educators' Perceptions of their School Environment, and School-Based Averages. This study took into consideration that Belize is a part of Central America and the Caribbean; the population is relatively small; the country is culturally and linguistically diverse; English is the official language; and Belize has a church-state system of public schools. The data-collection instrument, the *Belize Educator Survey*, was designed to give all educators a chance to contribute to the profile of their school directly. Either the Weighted-Indicator Scores Protocol or the Multilevel Achievement Estimation Protocol combined with the *Belize Educator Survey* would be feasible for estimating School-Level Achievement in Belize.

REFERENCES

AERA, APA, NCME. (2014). *Standards for educational and psychological testing.* Washington D.C.: AERA /APA/NCME.

Agasisti, T., Avvisati, F., & Longobardi, S. (2018). *Academic resilience: What schools and countries do to help disadvantaged students succeed in PISA.* Paris: OECD . Retrieved from http://dx.doi.org/10.1787/e22490ac-en

Airasian, P. W., & Madaus, G. F. (1983). Linking testing and instruction: Policy issues. *Journal of Educational Measurement,, 20*(2), 103-118.

Airasian, P. W., & Miranda, H. (2002). The role of assessments in the revised taxonomy. *Theory into Practice, 41*(4), 249-254.

Akaguri, L. (2013). Fee-free public or low-fee private basic education in rural Ghana: How does the cost influence the choice of the poor? *Compare: A Journal of Comparative and International Education, 44*(2), 140-161.

Amrein, A., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(18).

Ananiadou, K., & Claro, M. (2009). 21st century skills and competences for new millennium learners in OECD countries. *OECD Education Working Papers, 41*. Retrieved from http://dx.doi.org/10.1787/218525261154

Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.

Anderson, S., Medrich, E., & Fowler, D. (2007). Which achievement gap? *Phi Delta Kappan, 88*(7), 547-550. Retrieved from http://www.jstor.org/stable/20442316

Arcia, G. (2016). *Efficiency of public expenditures in education and health in Belize, 2003 - 2013.* Inter-American Development Bank. Retrieved from https://publications.iadb.org/publications/english/document/Efficiency-of-Public-Expenditures-in-Education-and-Health-in-Belize-2003---2013.pdf

Baker, W., Sammons, P., Siraj-Blatchford, I., Sylva, K., Melhuish, E., & Taggart, B. (2014). Aspirations, education and inequality in England: Insights for the effective provision of pre-school, primary and secondary education project. *Oxford Review of Education, 40*(5), 525-542. doi:DOI: 10.1080/03054985.2014.953921

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122-147.

Bandura, A. (1993). Percieved self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*(2), 117-148.

Bandura, A. (1995). Exercise of personal and collective efficacy in changing societies. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 1-13). Cambridge, UK: Cambridge university press.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In *Self-efficacy beliefs of adolescents* (pp. 307-337). Information Age Publishing.

Bandura, A. (2014). Bandura's instrument: Teacher self-efficacy scale. Retrieved from https://cpb-us-w2.wpmucdn.com/u.osu.edu/dist/2/5604/files/2014/09/Bandura-Instr-1sdm5sg.pdf

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Multifaceted impact of self-efficacy beliefs on academic functioning. *Child development, 67*(3), 1206-1222.

Banicky, L. A. (2000). Opportunity to learn. *Education Policy Brief, 7*.

Belize Ministry of Education. (2007). *Education statistical digest.* Ministry of Education. Belmopan: Belize Ministry of Education Youth Sports and Culture. Retrieved from http://www.moe.gov.bz/#37-education-statistical-digest

Belize Ministry of Education. (2012a). *Education rules: 2012 ammendment.* Retrieved from Government of Belize - Ministry of Education Youth Sports and Culture: http://www.moe.gov.bz/resources/education-legislation/

Belize Ministry of Education. (2012b). *Belize teaching service commission handbook.* Ministry of Education Youth Sports and Culture. Belmopan: Government of Belize. Retrieved from http://moe.gov.bz/tsc/

Belize Ministry of Education. (2012c). *Education Statistical Digest 2011-2012.* Ministry of Education. Belmopan: Belize Ministry of Education Youth Sports and Culture. Retrieved from http://www.moe.gov.bz/#

Belize Ministry of Education. (2018, June). Primary school examination preliminary results 2018. *Press Release*. Belmpoan, Belize: Government Press Office.

Bennett, A. J. (2008). *Education in Belize: A historical perspective.* Belize: The Angelus Press Ltd.

Binkley, M. E., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & & Rumble, M. (2012). Defining twenty-first century skills. In *In Assessment and teaching of 21st century skills* (pp. 17-66). Netherlands: Springer.

Bolland, N. O. (1997). *Struggles for freedom: Essays ona slavery, colonialism and culture in the Caribbean and Central America.* New Hope, Jamaica: Ian Randle Publishers Ltd.

Briggs, D., & Dominique, B. (2011). Due diligence and the evaluation of teachers: A review of the value-added analysis underlying the effectiveness rankings of los angeles unified school district teachers by the los angeles times. *National Education Policy Center*. Retrieved March 06, 2018, from http://nepc.colorado.edu/publication/due-diligence

Bruthiaux, P. (2003). Squaring the circles: Issues in modeling English worldwide. *International Journal of Applied Linguistics, 13*(2), 159-178. Retrieved from https://onlinelibrary.wiley.com/doi/pdf/10.1111/1473-4192.00042

Burke, M. A., & Sass, T. R. (2013). Classroom peer effects and student achievement. *Journal of Labour Economics, 31*(1), 51-82.

Burns, D. (2019). *Random number weighted probability*. Retrieved from ExcelJet: https://exceljet.net/formula/random-number-weighted-probability

Caprara, G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Journal of school psychology. *Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level, 44*(6), 473-490.

Cattell, R. B., & Voglemann, S. (1977). A comprehensive trial of the scree and KG criteria for determining the number of factors. *Multivariate Behavior Research, 12*(3), 289-325.

Chaudry, A., & Wimer, C. (2016). Poverty is not just an indicator: The relationship between income, poverty and child well-being. *Academic Pediatrics, 16*(3), 23-29.

Chin, M. M., & Chow, B. W. (2015). Classmate characteristics and student achievement in 33 countries: Classmate's past achievement, family socioeconomic status, educational resources, and attitude toward reading. *Journal of Educational Psychology, 170*(1), 152-169.

Cole, N. (1990). Conceptions of educational achievement. *Educational Researcher,* 2-7.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika1951, 16*, 297–334. doi:https://doi.org/10.1007/BF02310555

Cummins, J. (2016). Reflections on Cummins (1980),"The Cross-Lingual Dimensions of Language Proficiency: Implications for Bilingual Education and the Optimal Age Issue". *TESOL Quarterly, 50*(4), 940-944.

Cummins, J. (2017). Literacies and Language Education. *BICS and CALP: Empirical and theoretical status of the distinction*, 59-71.

Davies, J., & Ryan, J. (2013). *Primary school prinicpals EU/UB banana belt primary education project: Report on final visit - evaluation exercise.*

Deal, T. D., & Peterson, K. D. (2011). *Shaping school culture fieldbook: The Jossey-Bass Education Series* (2 ed.). San Francisco, CA: Jossey-Bass.

Deal, T. E., & Peterson, K. (2016). *Shaping school culture.* John Wiley & Sons.

Deal, T. E., & Peterson, K. D. (1998). How leaders influence culture of schools. *Education Leadership, 56*(1), 28-31.

Deal, T. E., & Peterson, K. D. (1999). Shaping school culture: The heart of leadership. *Adolescence, 34*(136), 802-.

Deal, T. E., & Peterson, K. D. (2009). *Shaping school culture: Pitfals, paradoxes, & Promises* (2 ed.). John Wiley & Sons.

Decker, K. (1994). Orthography development for Belize creole. *1994 MALC* (pp. 351-362). Mid-America Linguistics Conference.

Dixson, D. D., Keltner, D., Worrell, F. C., & Mello, Z. (2018). The magic of hope: Hope mediates the relationship between socioeconomic status and academic achievement. *Journal of Educational Research, 111*(4), 507-515.

Dolowitz, D., & David, M. (1996). Who learns what from whom: A review of the policy transfer literature. *Political Studies, XLIV,* 343-357. Retrieved from https://doi.org/10.1111/j.1467-9248.1996.tb00334.x

Duncan, G. J., & Magnuson, K. (2012). Socioeconomic status and cognitive functioning: Moving from correlation to causation. *Wiley Interdisciplinary Reviews: Cognitive Science, 3*(3), 377-386.

Elliott, S., & Bartlett, B. (2016). Opportunity to learn. *Psychology, School Psychology Online Publication*.

Erickson, F. (1987). Transformation and school success: The politics and culture of educational achievement. *Anthropology & Education Quarterly, 18*(4), 335-356.

Evans, R. (2005). Reframing the achievement gap. *The Phi Delta Kappan, 86*(8), 582-589. Retrieved from http://www.justor.org/stable/20441855

Faber, P. (2010). *Minister of education and youth address for the launch of the teacher training programme banana belt.* Belmopan: Belize Ministry of Education.

Farah, M. J., & Hackman, D. A. (2016). Chapter 17: SES, childhood experience and the neural bases of cognition. In *Oxford handbook of poverty and child development* (pp. 307-318).

Field, A. P. (2009). *Discovering statistics using SPSS: And sex and drugs and rock'n' roll* (3rd ed.). London: Sage.

Fullan, M., Rincon-Gallardo, S., & Hargreaves, A. (2015). Professional capital as accountability. *Education Policy Analysis Archives, 23*(15), 1-22. Retrieved from http://dx.doi.org/10.14507/epaa.v23.1998

Geiser, S., & Santelices, M. V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes.* University of California, Berkley. Retrieved from https://escholarship.org/uc/item/7306z0zf

Geisinger, K. F. (2000). Psychological testing at the end of the millennium: A brief historical review. *Professional Psychology: Research and Practice, 31*, 117–118.

Glen, S. (2015, September 21). *Variance inflation factor*. Retrieved from Statistics How To: https://www.statisticshowto.datasciencecentral.com/variance-inflation-factor

Glen, S. (2016, June 20). *Durbin watson test & test statistic*. Retrieved February 2020, from Statistics How To: https://www.statisticshowto.datasciencecentral.com/

Government of Belize. (2003). *Belize Education Act chapter 36: revised edition.* Belmopan: Government of Belize . Retrieved from http://www.moe.gov.bz/rubberdoc/779f108b0e8fa1149983f80d0757b32d.pdf

Government of Belize. (2011). *Belize's constitution of 1981 with ammendments through 2011*. Retrieved January 17, 2018, from Constituteproject.org: https://www.constituteproject.org/constitution/Belize_2011.pdf?lang=en

Government of Belize. (2014). *Official press release: Primary school examination preliminary results 2014.* Belize: Ministry of Education, Youth and Sports Press Release.

Grantham-McGregor, A., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. (2007). Developmental potential in the first 5 years for children in developing countries. *Education Policy and Evaluation Research, 369* , 60-70. Retrieved from www.thelancet.com

Gregersen, E. (2019). *Professor Leonard Klienrock*. (Encyclopædia Britannica, inc.) Retrieved 2020, from Encyclopædia Britannica: https://www.britannica.com/biography/Leonard-Kleinrock

Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher, 20*(5), 2-7. Retrieved 6 5, 2018, from http://journals.sagepub.com/doi/pdf/10.3102/0013189X020005002

Hanson, T. (2012). Measurement analysis of CHKS core and S3 module items. In M. O'Malley, A. Voight, T. L. Renshaw, & K. Eklund, *School climate, family structure, and academic achievement: A study of moderation effects* (p. 6). School Psychology Quarterly.

Härmä, J. (2016). School choice in rural Nigeria? The limits of low-fee private schooling in Kwara State. *Comparative Education, 52*(2), 246-266. Retrieved from https://doi.org/10.1080/03050068.2016.1142737

Haug, S. W. (1998). Ethnicity and etnically "mixed" identity in Belize: A study of primary school-age children. *Anthropology & Education Quarterly*, *29*(1), 44-67.

Heafner, T. L., & Fitchett, P. G. (2015). An opportunity to learn US history: What NAEP data suggest regarding the opportunity gap. *The High School Journal, 98*(3), 226-249.

Heyneman, S. P., & Loxley, W. A. (1983). The effect of primary-school quality on academic achievemetn across twenty-nine high- and low-income countries. *American Journal of Sociology*, 1162-1194.

Hodara, M., & Cox, M. (2016). *Developmental education and college readiness at the University of Alaska. REL 2016-123. Regional Educational Laboratory Northwest.* Department of Education, Institute of of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northwest. Retrieved from http://ies.ed.gov/ncee/edlabs.

Hodge, M., & Welch, J. (2016). An implementation perspective: Relevant lessons from no child left behind (NCLB) for the implementation of every student succeeds act (ESSA). *Journal of Ethical Educational Leadership, 3*(9), 1-17.

Holbein, J. B., & Ladd, H. F. (2017). Accountability pressure: Regression discontinuity estimates of how No Child Left Behind influenced student behavior. *Economics of Education Review*, *58*, 55-67.

Hox, J. J., Maas, C. J., & Brinkhuis, M. J. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica, 64*(2), 157-170. doi:10.1111/j.1467-9574.2009.00445.x

Human Rights Watch. (2014, November 17). *25th anniversary of the convention on the rights of the child*. Retrieved from Human Rights Watch: https://www.hrw.org/news/2014/11/17/25th-anniversary-convention-rights-child

Humes, A. (2008, October 10). Can we talk? QUADS truancy unit to train school wardens and truancy officers in the art of effective communications. *Amandala Press*. Retrieved from http://amandala.com.bz/news/can-we-talk-quads-truancy-unit-to-train-school-wardens-and-truancy-officers-in-the-art-of-effective-communication/

Humes, A. (2016, June 10). PSE 2016 results released; grades stay average. *Breaking Belize News*. Retrieved from https://www.breakingbelizenews.com/2016/06/10/pse-2016-results-released-grades-stay-average/

Hurtado, S. (2001). Linking diversity and educatinal purpose: How diversity affects the classroom environment and student development. In G. Orfield, *Diversity challenged: Evidence on the impact of affirmative action* (pp. 187-203). Cambridge: Harvard Education Publishing Group.

IBM Corporation. (2016). IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.

IBM Corporation. (2019, September 29). *KMO & Bartlett's Test*. Retrieved from IBM Knowledge Center:

https://www.ibm.com/support/knowledgecenter/SSLVMB_26.0.0/statistics_casestudies_project_ddita/spss/tutorials/fac_telco_kmo_01.html

International Test Commission. (2001). ITC Guidelines on test use. *International Journal of Testing, 1*(2), 93-114. Retrieved 6 5, 2018, from https://www.intestcom.org/files/guideline_test_use.pdf

Jerald, C. D. (2009). Defining a 21st century education. *The Center for Public Education, 16*.

Jhurree, V. (2005). Technology integration in education in developing countries: Guidelines to policy makers. *International Education Journal, 6*(4), 467-483. Retrieved from http://iej.chb.net

Judge, T. A., & Bono, J. E. (2001). Journal of applied Psychology. *Relationship of core self-evaluations traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability—with job satisfaction and job performance: A meta-analysis, 86*(1), 80-92.

Knaub, J. R. (2011). Finite population correction (FPC) factor. In P. J. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods* (pp. 285-286). SAGE Publications, Ltd.

Kolb, A., & Kolb, D. A. (2005, June). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning & Education, 4*(2), 193-212. Retrieved from http://www.jstor.org/stable/40214287 Accessed: 11-01-2018 01:03 UTC

Lavis, J. N., Posada, F. B., Haines, A., & Osei, E. (2004). Use of research to inform public policymaking. *The Lancet, 364*(9445), 1615-1621. Retrieved from https://doi.org/10.1016/S0140-6736(04)17317-0

Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research, and Evaluation, 12*(2), 1-11.

LeTendre, G. (2002). Advancements in conceptualizing and analyzing cultural effects in cross-national studies of educational achievement. In A. C. Porter, & A. Gamoran, *Methodological advances in cross-national surveys of educational achievement* (pp. 198-228). National Academies Press. Retrieved from http://www.nap.edu/catalog/10322.html

Linn, R. L. (2002). The measurement of student achievement in international studies. In A. C. Porter, & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp. 27-57). National Research Council.

Martin, J. P. (2001). *Starting strong: Early childhood education and care - Introductory presentation.* Stockholm: OECD.

Masinoa, S., & Niño-Zarazúa, M. (2016). What works to improve the quality of student learning in developing countries? *International Journal of Educational Development, 48*, 53-65. Retrieved 01 15, 2018, from https://www.sciencedirect.com/science/article/pii/S0738059315300146

Massachusetts Department of Elementary and Secondary Education. (2018, July 19). *Educator Licensure and Renewal*. Retrieved from Mass.Gov: https://gateway.edu.state.ma.us/elar/licensurehelp/LicenseRequirementsPageControl.ser?COMMAND.back

Matsueda, R. L. (2011, June 25). Key advances in the history of structural equation modeling. *Working Paper no. 114*. Center for Statistics and the Social Sciences, University of Washington.

McCoy, D. C., & Zuilkowski, S. S. (2015). Poverty, Physical stature, and cognitive skills: Mechanisms underlying children's school enrollment in Zambia. *Developmental Psychology, 51*(5), 600–614.

McKibben, J. N. (2004). Racial and ethnic composition. In J. S. Siegel, & D. A. Swanson (Eds.), *The methods and materials of demography* (pp. 175-189). California: Elsevier Academic Press.

Moss, P. A., Pullin, D., Gee, J. P., & Haertel, E. H. (2005). The idea of testing: Psychometric and sociocultural perspectives. *Measurement, 3*(2), 63-83.

Murphy, D. (2012). *Value in value-added.* Pearson Education, Inc. Retrieved 7/26/2017

Muthen, B., & Muthen, L. (1998-2019). Mplus 8.3. *8.3*. California. Retrieved from www.StatModel.com

Muthén, B., Huang, L. C., Khoo, S. T., Goff, G. N., Novak, J. R., & Shih, J. C. (1995). Opportunity to Learn effects on achievement: Analytical aspects. *Educational Evaluation and Policy Analysis, 17*(3), 371-403.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide* (Eighth ed.). Los Angeles, California: Muthén & Muthén.

Näslund-Hadley, E., Alonzo, H., & Martin, D. (2013). *Challenges and opportunities in the Belize education sector.* Inter-American Development Bank, Education Division. Inter-American Development Bank. Retrieved from http://www.iadb.org

NCES. (2017, December 11). *International Activities Program*. Retrieved from National Center for Education Statistics: https://nces.ed.gov/surveys/international/about.asp

Nebraska Department of Education. (2018, July 19). *Nebraska Department of Education*. Retrieved from Teacher Certification: https://dc2.education.ne.gov/tc_interactive_teaching2/TeachingInitialCertficate.aspx

Nelson, L. R. (2016). Some observations on the scree test, and on coefficient alpha. *Research Methodology and Cognitive Science, 3*(1), 1-17.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*(3), 396-402.

OECD. (2001). *Starting strong: Early childhood education and care.* Paris: OECD. Retrieved from www.sourceOECD.org

OECD. (2009). *Doing better for children* . Paris: OECD.

OECD. (2011). *Against the odds: Disadvantaged students who succeed in school.* Paris: OECD Publishing.

O'Malley, M., Voight, A., Renshaw, T., & Eklund, K. (2014). School climate, family structure, and academic achievement: A study of moderation effects. *School Psychology Quartely, 30*(1), 1-16.

Opdenakker, M.-C., & Van Damme, J. (2007). Do school context, student composition and school leadership affect school practice and outcomes in secondary education? *British Educational Research Journal, 33*(2), 179-206. Retrieved from http://www.jstor.org/stable/30032740

P21®. (2018, November). *P21 Vision and Mission*. Retrieved from P21 Partnership for 21st Century Learning: http://www.p21.org/about-us/our-mission

Paufler, N., & Amrein-Beardsley, A. (2014). The random assignment of students into elementary classrooms: Implications for value-added analyses and interpretations. *American Educational Research Journal, 51*(2), 328-362. doi:10.3102/0002831213508299

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measureent, and improvement of classroom processes: Standardized observaton can leverage capacity. *Educational Researcher, 38*(2), 109-119.

Plomp, T. (1990). IEA: Its role and plans for international comparative research in education. *American Educational Research Association.* Boston.

Policy Planning Research & Evaluation Unit. (2015). *Abstract of education statistics 2014/2015.* Belize: Ministry of Education, Youth, & Sports.

Pritchard, I. (1995). *Judging Standards in Standards-Based Reform.* http://www.ecs.org: Education Commission of the States.

Purves, A. (1987). IEA agenda for the future. *International Review of Education, 33*, 103-107.

Raudenbush, S. W., Rowan, B., & Cheong, Y. F. (1992). Contextual effects on the self-perceived efficacy of high school teachers. *Sociology of Education, 65*(2), 150-167.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4*(3), 207-230.

Ross, J. A. (1992). Teacher efficacy and the effects of coaching on student achievement. *Canadian Journal of Education/Revue canadienne de l'education, 17*(1), 51-65.

Samuelsson, I. (2011). Why we should begin early with ESD: The role of early childhood education. *Internation Journal of Early Childhood (IJEC), 43*, 103-118. doi:10.1007/s13158-011-0034-x

San Pedro Sun. (2017, June 6). Belize district tops 2017 primary school examinations. *The San Pedro Sun*. Retrieved from https://www.sanpedrosun.com/education/2017/06/06/belize-district-tops-2017-primary-school-examinations/

Scholtus, S., van de Laar, R., & Willenborg, L. (2014). *The memobust handbook on methodology for modern business statistics.* Retrieved from https://ec.europa.eu/eurostat/cros/system/files/Imputation-01-T-Main%20Module%20v1.0_0.pdf

Schuh-Moore, A., DeStefano, J., & Adelman, E. (2012). Opportunity to Learn: A high impact strategy for improving educational outcomes in developing countries.

Seaman, S., White, I. R., Copas, A. J., & Li, L. (2012). Combining multiple imputation and inverse-probability weighting. *Biometrics, 68*(1), 129–137. doi:doi: 10.1111/j.1541-0420.2011.01666.x

Segretin, M. S., Hermida, M. J., Fracchia, C. S., Ruetti, E., & Lipina, S. J. (2016). Childhood poverty and cognitive development in Latin America in the 21st century. *Child and adolescent development in Latin America*(152), 9-29. doi:10.1002/cad.20162

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, L. E., & Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24*(1), 19-35. Retrieved from https://journals.sagepub.com/doi/pdf/10.1177/0734282905285237?casa_token=3ivIp5q2HuIAAAAA%3A2hp-W4HJ7m79Lr8Su-vehzyOdkL7wbfK-AbXLCRdYWbaRj-CiDX7JhWrlSTT_JrFSgRFZTj9KxFS

Sharkins, K. A., Leger, S. E., & Ernest, J. M. (2017). Examining effects of poverty, maternal depression and children's self-regulation abilities on the develpment of language and cognition in early childhood: A head start perspective. *Early Childhood Education Journal, 45*, 493-498.

Shoman, A. (1994). *A history of Belize in 13 chapters.* Belize City: Angelus Press Ltd.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of educational psychology, 85*(4), 571-581.

Smith, M. (2002). Drawing inferences for national policy from large-scale cross-national education surveys. In NationalResearchCouncil, *Methodological advances in cross-national surveys of educational achievement* (pp. 295-317). Washington, DC: National Academies Press. Retrieved from https://doi.org/10.17226/10322.

Statistical Institute of Belize. (2013). *Belize population and housing census report 2010.* Belmopan: Statistical Institute of Belize. Retrieved from www.statisticsbelize.org.bz

Stephens, M., Warren, L. K., & Harner, A. L. (2015). *Comparative indicators of education in the United States and other G-20 countries:2015 (NCES2016-100).* U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Taguma, M., Litjens, I., & Makowiecki, K. (2012). Quality Matters in Early Childhood Education and care: Finland. *OECD*.

The World Bank. (2018). *World Bank Country and Lending Groups*. Retrieved January 27, 2018, from World Bank Group: Data & Statistics: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

Tooley, J., & Dixon, P. (2006). 'De facto' privatisation of education and the poor: Implications of a study from sub-Saharan Africa and India. *Compare: A Journal of Comparative and International Education, 36*(4), 443-462. Retrieved from https://doi.org/10.1080/03057920601024891

Tran, M. T., Young, R. L., & Di Lella, J. D. (1994). Multicultural education courses and the student teacher: Eliminating stereotypical attitudes inour ethnically diverse classroom. *Journal of Teacher Education, 45*(3), 183-189.

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times.* Wiley & Sons.

UNESCO. (2006). *Teachers and educational quality: Monitoring global needs for 2015.* Montreal: UNESCO Institute for Statistics.

UNESCO. (2007). Belize: Caribbean symposium on inclusive education. *Caribbean Symposium on 'Inclusive Education'* (pp. 1-8). Kingston, Jamaica: UNESCO International Bureau of Education.

Usher-Tate, B. (2014). *Efficacy and perceptions of educators preparing students for the national examinations in Belize.* unpublished paper (graduate coursework).

Usher-Tate, B. (2015). Educator efficacy in southern Belize. *Poster Presentation: American Educational Research Association*. Chicago, IL.

Valadez, L. (2010). Medium-term effects of household poverty on child well-being; A study in a subsample of pre-school children in rural Mexico. *Journal of International Development*, *22*, 1146-1161.

Vockley, M. (2007). Maximizing the impact: The pivotal role of technology in a 21st century education system.

Waddar, M. S., & Aminabhavi, V. A. (2010). Self-efficacy and emotional intelligence of PG students. *Journal of the Indian Academy of Applied Psychology, 36*(2), 339-345.

Wainer, H., & Thissen, D. (1996). How is reliability related to the quality of test scores? What is the effect of local dependence on reliabilty? *Education Measurement: Issues and Practice*, *15*(1), 22-29.

Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness* . New Teacher Project.

Zopluoglu, C., & Davenport, E. C. (2017). A note on using eigenvalues in dimensionality assessment. *Research and Evaluation, 22*(7), 1-10.

Zumbach, D. (2010). *The 'Heyneman-Loxley Effect' in Sub-Saharan Africa: School quality, socioeconomic status and national economic development.* Retrieved 2019, from https://ssrn.com/abstract=1815463 or http://dx.doi.org/10.2139/ssrn.1815463

APPENDIX A

Index of Abbreviations Used in this Study

CTX The estimation variable for School Context

ECEC Early Childhood Education Center

EEB The estimation variable for the latent variable identified as the Educator Effect / Bias

EFF The estimation variable for Educators-Efficacy

G-20 Group of 20 developed countries participating in OECD & IEA studies: Argentina, Australia, Brazil, Canada, China, France, Germany, Indonesia, Italy, Japan, Mexico, Republic of Korea, Russian Federation, Saudi Arabia, South Africa, Turkey, United Kingdom, United States. https://nces.ed.gov/pubs2016/2016100.pdf

IDB International Development Bank

IEA The International Association for the Evaluation of Educational Achievement was founded in 1967. Independent collaboration between research institutions and agencies across member countries.

MOE The Belize Ministry of Education

MAEP Multilevel Achievement Estimation Protocol

OECD Organization for Economic Co-operation and Development

OTL Opportunity to Learn reference to the concept and the estimation variable

PIRLS Progress in International Reading Literacy Study - This is an IEA study of reading achievement in fourth graders https://nces.ed.gov/surveys/pirls/

PISA Programme for International Student Assessment - This is a triennial international survey which aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students. http://www.oecd.org/pisa/aboutpisa/

PRC The estimation variable for the Educators' Perceptions of Their School Environment

PSE Primary School Examinations (7-section national standardized exams given to 8th grade students and administered across two whole-days)

SBA School-Based Averages is the estimation variable for educator supplied class averages in four core subjects: English, Mathematics, Science, and Social Studies

RIG The variable for overall School-Rigor variable

TALIS Teaching and Learning International Survey http://www.oecd.org/edu/school/talis.htm

TIMSS Trends in International Mathematics and Science Study - Since 1995 (in 4-year increments), this IEA study assesses the Mathematics and Science knowledge of students in grades 4 and 8 in different countries. The 2015 study was extended to grade 12 students. https://nces.ed.gov/timss/

WISP Weighted-Indicator Scores Protocol

APPENDIX B

Internal Consistency (Reliability)

Estimates for Internal Consistency

Estimation Variable	Coefficient alpha (α)	Number of items	Sample size
2014 Educator Survey	.876	35	142
2018 The *Belize Educator Survey* Pilot			20
Efficacy	.909	18	
Learning Environment	.864	6	
Influence	.748	6	
Professionalism	.801	6	
Perceptions of Their School Environment (PRC)	.800	27	
Curriculum & Assessment	.663	12	
School Culture & Climate	.619	8	
Engagement	.806	7	
Opportunity to Learn : Access (OTL)	.574	7	
Opportunity to Learn : Resources (CTX)	.605	3	
School-Based Averages (SBA)	.939	4	

APPENDIX C:

Appendix C1: Permission documents (Government of Belize and IRB)

Ministry of Education, Youth, Sports & Culture

GEN/21/18(150) VOL. IV

November 14, 2018

TO WHOM IT MAY CONCERN

I write in support of the research that **Ms. Betty Jean Usher-Tate** proposes to conduct as part of her doctoral dissertation project.

I hereby grant permission to Ms. Usher-Tate to conduct her research on "**Estimating School-level Educational Achievement in Belize**" under the following protocols.

1. The data provided shall not be changed or tampered with in any way.
2. Access to these data shall be for the purposes of analysis, planning, and conducting research.
3. The information is not to be misused in any way.
4. Confidentiality of any personal information must be maintained.
5. No confidential information is to be divulged to anyone who is not authorized to access this information.
6. Any information derived from the data which may be published or generally circulated shall not identify or reveal the confidential information of any individual or schools.

Kindly extend every courtesy to Ms. Betty Jean Usher-Tate in this endeavour.

Sincerely,

Carol Babb

Carol Babb, Ed. D.
Chief Education Officer

West Block, City of Belmopan
Belize
Central America

Tel: (501) 822-2329/2380,0385
Fax: (501) 822-3389
E-mail: moeducation@moes.gov.bz

Nebraska University of Lincoln

Official Approval Letter for IRB project #18717 - New Project Form

January 7, 2019 - official approval letter

Betty Jean Usher-Tate
IANR/CEHS

Kurt Geisinger
BUROS
TEAC 21G, UNL, 685880345

IRB Number: 20190118717 EX
Project ID: 18717
Project Title: 2018 Belize Educator Survey

Dear Betty Jean:

This letter is to officially notify you of the certification of exemption of your project for the Protection of Human Subjects. Your proposal is in compliance with this institution's Federal Wide Assurance 00002258 and the DHHS Regulations for the Protection of Human Subjects (45 CFR 46) and has been classified as exempt. Exempt categories are listed within HRPP Policy #4.001: Exempt Research available at: http://research.unl.edu/researchcompliance/policies-procedures/.

o Date of Final Exemption: 01/07/2019
o Review conducted using exempt category 2 at 45 CFR 46.101
o Funding (Grant congruency, OSP Project/Form ID and Funding Sponsor Award Number, if applicable): N/A

We wish to remind you that the principal investigator is responsible for reporting to this Board any of the following events within 48 hours of the event:
* Any serious event (including on-site and off-site adverse events, injuries, side effects, deaths, or other problems) which in the opinion of the local investigator was unanticipated, involved risk to subjects or others, and was possibly related to the research procedures;
* Any serious accidental or unintentional change to the IRB-approved protocol that involves risk or has the potential to recur;
* Any publication in the literature, safety monitoring report, interim result or other finding that indicates an unexpected change to the risk/benefit ratio of the research;
* Any breach in confidentiality or compromise in data privacy related to the subject or others; or
* Any complaint of a subject that indicates an unanticipated risk or that cannot be resolved by the research staff.

This project should be conducted in full accordance with all applicable sections of the IRB Guidelines and you should notify the IRB immediately of any proposed changes that may affect the exempt status of your research project. You should report any unanticipated problems involving risks to the participants or others to the Board.

If you have any questions, please contact the IRB office at 402-472-6965.

Sincerely,

Becky R. Freeman, CIP
for the IRB

Appendix C2: Flyer

NOTICE OF UPCOMING STUDY:

The Belize Educator Survey

Lead Researcher: Betty-Jean Usher-Tate, is a fellow Belizean educator. She is in the dissertation phase of a Ph.D. in Quantitative Qualitative and Psychometric Methods (QQPM) and Educational Administration (EDAD) at the University of Nebraska in Lincoln.

Purpose: *The Belize Educator Survey* is one part of her dissertation research project, to estimate school-level educational achievement in Belize. The outcome of the study could potentially serve as a preliminary model designed specifically for the context of our education system.

Data Collection: *The Belize Educator Survey* takes approximately 25-30 minutes to complete. All educator contributions are confidential and will inform the collective voice of the school. The results of this survey will also provide information that will be used to evaluate the psychometrics of Betty Jean's proposed conceptual model for school-level achievement. Unlike other models, this model emphasizes input from educators in addition to other academic performance measures.

Participation (voluntary): In recognition that educators at every level make an impact on the school as a whole, I am seeking input from every educator associated with the school. Please encourage your fellow educators to complete the survey and be a part of the collective voice that builds a comprehensive profile of your school.

Advance preparation: Teachers could calculate their class averages in Math, English, Science and Social Studies for the class(es) they taught in the 2017-2018 academic year.

Incentives: An opportunity to win a handmade Arabian satchel or one of two *Kindle* tablets and also knowing that you contributed to having the true context of your school represented meaningfully in the statistics.

Appendix C3: Informed Consent

University of Nebraska Lincoln

COLLEGE OF EDUCATION AND HUMAN SCIENCES
Educational Psychology

Informed Consent Notice
Project ID: 51290 **IRB#**: **18717**

Title: 2018 Educator Survey

Purpose: This survey is a part of a dissertation research project. This survey gives educators, from all levels in a school, an opportunity to contribute and define a holistic school profile beyond performance the Standard-6 examinations.

Procedures: You will be asked to respond to items in this survey (approximately 30 minutes). The information from this survey will be combined with other data to estimate school-level educational achievement in Stann Creek District, which is the overarching purpose of the dissertation study. Copies of the dissertation will be shared with Belize's Ministry of Education, and parts thereof might be included in presentations at international conferences or articles in professional journals.

Benefits: Your contribution (voice) may serve to inform practice and influence future decisions intended to improve educational services and/or policies at your school and other schools in the country.

Risks/Discomforts: There are no known risks or discomforts associated with this research.

Confidentiality: Survey responses will remain confidential. In the dissertation and any subsequent publications, due care will be taken to filter out or de-identify peculiar or sensitive information that could be linked to a particular individual or school.

Compensation: All educators who provide a viable contact information on the ticket included in the packet will be entered in a drawing to win either a unisex handmade Arabian Satchel or one of two Kindle Fire 7 tablets. Participation is voluntary. Completion of the survey is not a requirement and the contact information will only be used for this raffle. The identity of each winner will not be published. The winners will be notified directly to make arrangements to collect their prize. There is no other monetary reward or gift for participating in this project. The overall odds of winning a Kindle Fire 7 tablet in the sweepstake is about 1 in 200.

Opportunity to Ask Questions: You may ask questions concerning this research (directly to the principal investigator or an intermediary of your choice). To voice concerns about the research or if you have any questions about your rights as a research participant, please contact the principal investigator or the University of Nebraska-Lincoln Institutional Review Board at (402) 472-6965.

Freedom to Withdraw: Participation in this study is voluntary. You can refuse to participate or withdraw at any time without harming your relationship with the researchers, the University of Nebraska-Lincoln, or education affiliations Belize.

Consent: You are voluntarily making a decision whether or not to participate in this research study. By marking a circle on the front page of the survey sheet that indicate participation options, you certify that you have read and understood the information presented here and that you freely decided to participate.

Name and email contact for the investigators:
Betty-Jean Usher-Tate, MA, *Principal Investigator* ushertate@huskers.unl.edu
Kurt Geisinger, Ph.D., *Professor* kgeisinger2@unl.edu

You may keep this sheet for future reference.

114 Teachers College Hall / P.O. 880345 / Lincoln, NE 68588-0345
http://edpsyc.unl.edu

Appendix C4: Survey Materials/Instructions

THE BELIZE EDUCATOR SURVEY

Contents of your envelope (5 pages)

- Notice for informed consent – [1 page]
- This page – [Instruction sheet with Sweepstake/Raffle ticket at the bottom]
- The Educator Survey (pilot) – [3 pages, each printed on both sides]

Instructions

READ carefully - Your participation status and your responses will be kept strictly confidential. I can assure confidentiality, especially because of the strict university rules I have to adhere to for conducting my dissertation study. I will never report information that can be linked back to any specific individual. The goal of this study is to answer *school-level* research questions, that can be used objectively to help the school- not to pass judgement on educators.

RESPOND - There are no right or wrong answers. Please respond to ALL items on the two sides of each survey sheet (approx. 30 minutes).

SWEEPSTAKE (raffle) - You are eligible to enter the sweepstake even if you did not complete the survey. (I personally covered the cost of the two Kindles offered in the sweepstake and one of my professors donated the unisex handmade Arabian satchel which he brought back from Saudi Arabia.)

SEAL THE ENVELOPE – Please put the survey and this page back in the envelope you were provided and SEAL IT – confidentiality is critically important for my study.

RETURN - Please give your *sealed envelope* to the person in charge of collecting them at your school. I really appreciate your participation.

SWEEPSTAKE ENTRY

Please put your contact information below. *Why?* So I can personally contact you and make arrangements for you to collect your Kindle if your entry is selected. *No public announcements.*

Name: ______________________________ ***and either your***

Phone number: ______________________ ***OR***

Email address: ______________________________

Appendix C5: The *Belize Educator Survey*

The Belize Educator Survey

This survey is part of the process to develop a model for estimating school-level educational achievement in Belize. Participants are generally expected to either shade circles or supply information in the empty spaces provided with a single word or number.

This survey is voluntary and confidentiality is assured. It typically takes 20 to 30 minutes to complete this survey. For my dissertation records, I am asking that you state your name and also indicate your participation status. This information will not be shared.

Name ____

Participation status: ◯ YES (I will participate) ◯ NO (I will not participate)

This survey is printed with items on both sides of each page.
It is very important to this study that you respond to items on each side.

	This first section of the survey has questions on a variety of issues. *Please use the spaces provided or shade circles to indicate your responses.*	
01	How many classes total did the school have? (from Infant I to Standard 6)	
02	How many students total were enrolled in the school in the 2017-2018 school year?	
03	How many years of experience did you have as an educator by the end of the 2017-2018 year?	
04	How long did you work / have you worked at the school you identified on page one?	
05	Approximately how many times were you absent in the 2017-2018 school year?	
06	Who took care of your personal responsibilities at school when you were absent? ◯ no one ◯ student ◯ parent/volunteer ◯ teacher ◯ principal/VP	
07	How did you normally get from home to school? ◯ Taxi/Regular Bus ◯ Bicycle ◯ Walk ◯ Private vehicle ◯ School Bus	
08	What was your role(s) at the school in 2017-2018? (*shade all that applied to you*) ◯ on Internship ◯ Assistant Teacher ◯ Teacher ◯ Vice Principal ◯ Principal	
09	Which of the following best describes how you feel about the behavior of the students you are responsible for? ◯ Not well behaved ◯ Not consistent ◯ Adequate ◯ Good ◯ Excellent	
10	Which of the following best describes the physical condition of your work environment (class/school)? ◯ Poor/unsafe ◯ Inadequate/cramped ◯ Adequate ◯ Good ◯ Excellent	
11	What is your highest level of formal schooling? ◯ High School ◯ 6^{th} form/Associate ◯ Bachelor ◯ Master ◯ Doctor	
12	If you had a bookshelf or reading resource area in your class/office, how many books did you have in the class? ◯ 0-25 ◯ 26-50 ◯ 51-75 ◯ 76-100 ◯ more than 100	
13	Approximately how many books total are in the house you live (all kinds - textbook, novel, cookbook, bible, *etc.*)? ◯ 0-25 ◯ 26-50 ◯ 51-75 ◯ 76-100 ◯ more than 100	
14	What do you identify as your personal ethnic group(s)? *(shade all that apply to you)* ◯ Asian ◯ Creole (Kriol) ◯ East Indian ◯ Garifuna ◯ Maya ◯ Mennonite ◯ Mestizo ◯ Other	
15	Which ethnic group or groups were represented among the students in your class/school? (*shade all that applied*) ◯ Asian ◯ Creole (Kriol) ◯ East Indian ◯ Garifuna ◯ Maya ◯ Mennonite ◯ Mestizo ◯ Other	
16	Which of the following language(s) can you speak? (*Please shade all that apply to you*) ◯ Chinese ◯ Creole (Kriol) ◯ English ◯ Garifuna ◯ Maya ◯ German ◯ Spanish ◯ Other	
17	Which of the following do you use for news and current events? *Please shade all that apply to you* ◯ News Paper ◯ Magazines ◯ Public Meetings ◯ Conversations ◯ Radio ◯ TV News ◯ SMS (texts) ◯ email ◯ Social Media (*FB, Twitter...*) ◯ Internet ◯ Google ◯ Wikipedia	

Page 2. *Please go on to the next page.*

18	In an average day, what percentage of your time at work did you spend disciplining students? ◯ 0% ◯ 10% ◯ 20% ◯ 30% ◯ 40% ◯ 50% ◯ 60% ◯ 70% ◯ 80% ◯ 90% ◯ 100%
19	What was the overall attendance rate for the students you were responsible for? (*shade your best-guess/estimate*) ◯ 0% ◯ 10% ◯ 20% ◯ 30% ◯ 40% ◯ 50% ◯ 60% ◯ 70% ◯ 80% ◯ 90% ◯ 100%
20	What was the overall attendance rate for teachers? (*shade your best-guess/estimate*) ◯ 0% ◯ 10% ◯ 20% ◯ 30% ◯ 40% ◯ 50% ◯ 60% ◯ 70% ◯ 80% ◯ 90% ◯ 100%
21	When a student(s) struggled to learn a concept or skill, what percent of those instances did you *make time to help*? ◯ 0% ◯ 10% ◯ 20% ◯ 30% ◯ 40% ◯ 50% ◯ 60% ◯ 70% ◯ 80% ◯ 90% ◯ 100%
22	In 2017-2018, what percentage of class time was typically spent with students engaged in learning tasks/activities? ◯ 0% ◯ 10% ◯ 20% ◯ 30% ◯ 40% ◯ 50% ◯ 60% ◯ 70% ◯ 80% ◯ 90% ◯ 100%
23	How much of the curriculum for English did you cover well/satisfactorily with your students? ◯ I did not teach ◯ 0% ◯ 10% ◯ 20% ◯ 30% ◯ 40% ◯ 50% ◯ 60% ◯ 70% ◯ 80% ◯ 90% ◯ 100%
24	How much of the curriculum for Math did you cover well/satisfactorily with your students? ◯ I did not teach ◯ 0% ◯ 10% ◯ 20% ◯ 30% ◯ 40% ◯ 50% ◯ 60% ◯ 70% ◯ 80% ◯ 90% ◯ 100%

25	Consider the school you worked during 2017-2018 when you read the passage below. By your estimation, what is the earliest class level where teachers would use short passages like this example? **<u>What Happened</u>** It was raining this morning. I got Alex to play with me in the house. He threw the baseball to me but I missed it. The ball knocked a picture frame off the wall. We covered the pieces of glass with newspaper. Mom won't be happy when she gets home. *Curriculum Details*: past tense narrative (story); simple & compound words and sentences. *Paragraph Structure*: 6 related sentences across 3 lines (50 words total) In other words, what is the earliest level that students at the school would be expected to read and understand the story-line of the passage above - *<u>What Happened</u>*?	Infant ① Infant ② Standard ① ② ③ ④ ⑤ ⑥
26	By your estimation, what was the earliest class level where two digit sums with carry-overs was taught in the school you were at in 2017-2018? **<u>Arithmetic:</u>** 19 + 23 ____ or 76 + 8 = ____ *Curriculum Details*: place value Page (units and tens); two digit sums (addition) In other words, what was the lowest class where you think the students were expected solve Arithmetic problems like the examples above?	Infant ① Infant ② Standard ① ② ③ ④ ⑤ ⑥

Please read each statement and indicate the level to which you disagree or agree with each statement.
For this section, think only about the school you worked in the 2017-2018 school year.

1____**10**

① = Strongly Disagree (negative impression) to Strongly Agree = ⑩ (positive impression)

01	As an educator, I felt physically safe at the school.	①②③④⑤⑥⑦⑧⑨⑩
02	Students at the school participated in Festival of Arts.	①②③④⑤⑥⑦⑧⑨⑩
03	Communication between administration and teachers was professional and effective.	①②③④⑤⑥⑦⑧⑨⑩
04	Teachers were excited about what they taught.	①②③④⑤⑥⑦⑧⑨⑩
05	Teachers used a variety of teaching strategies to help students develop skills and learn.	①②③④⑤⑥⑦⑧⑨⑩
06	Students felt in charge of their own learning.	①②③④⑤⑥⑦⑧⑨⑩
07	Teachers provided students with challenging learning experiences.	①②③④⑤⑥⑦⑧⑨⑩
08	Students in your school participated in sporting competitions with other schools.	①②③④⑤⑥⑦⑧⑨⑩
09	Teachers frequently gave students feedback about classwork & homework.	①②③④⑤⑥⑦⑧⑨⑩
10	Students completed assignments and studied without reminders.	①②③④⑤⑥⑦⑧⑨⑩
11	The expectations for appropriate student behavior were similar from teacher to teacher.	①②③④⑤⑥⑦⑧⑨⑩
12	Teachers clearly communicated what they expected students to learn.	①②③④⑤⑥⑦⑧⑨⑩
13	Teachers emphasized the importance of effort.	①②③④⑤⑥⑦⑧⑨⑩
14	Classroom assessments were aligned with learning expectations.	①②③④⑤⑥⑦⑧⑨⑩
15	There were enough opportunities for extracurricular activities for students.	①②③④⑤⑥⑦⑧⑨⑩
16	There was mutual respect for cultural differences among staff.	①②③④⑤⑥⑦⑧⑨⑩
17	Teachers modified their instruction based on observations of participation/interest.	①②③④⑤⑥⑦⑧⑨⑩
18	Students participated in or supported extracurricular activities planned by the school.	①②③④⑤⑥⑦⑧⑨⑩
19	The school rules and discipline policies were applied fairly to all students.	①②③④⑤⑥⑦⑧⑨⑩
20	Classroom activities often required students to reflect on their own thinking.	①②③④⑤⑥⑦⑧⑨⑩
21	Teachers monitored students' learning & adjusted instruction to suit the situation.	①②③④⑤⑥⑦⑧⑨⑩
22	Students of all cultural backgrounds had a voice at the school.	①②③④⑤⑥⑦⑧⑨⑩
23	Students got opportunities to express themselves in Art (music/drama/craft).	①②③④⑤⑥⑦⑧⑨⑩
24	Behavior and learning expectations were clearly explained to students.	①②③④⑤⑥⑦⑧⑨⑩
25	Students in this school seem motivated to do well in school (and PSE for Std. 6).	①②③④⑤⑥⑦⑧⑨⑩
26	Teachers checked students' understanding/learning using a variety of assessments.	①②③④⑤⑥⑦⑧⑨⑩
27	At this school, teachers & administrators believed all students could be successful.	①②③④⑤⑥⑦⑧⑨⑩
28	Administrators emphasized the importance of effort.	①②③④⑤⑥⑦⑧⑨⑩
29	There was mutual respect for cultural differences among students.	①②③④⑤⑥⑦⑧⑨⑩

Page 4. *Please go on to the next page.*

Please consider each item in this section and respond based on your own belief or your own level of confidence about your capacity to accomplish that specific tasks/goals.

This is about you personally as an educator.

1 ____ **10**

① = lowest /weakest level (negative impression) — to — highest/strongest level = ⑩ (positive impression)

01	How much can you do to make school a safe place?	①②③④⑤⑥⑦⑧⑨⑩
02	How much are you able to contribute to curriculum planning & development?	①②③④⑤⑥⑦⑧⑨⑩
03	How much can you do to get students to trust/respect educators?	①②③④⑤⑥⑦⑧⑨⑩
04	How much can you influence decisions to offer students extra classes?	①②③④⑤⑥⑦⑧⑨⑩
05	How much can you do to control disruptive behavior (classroom & playground)?	①②③④⑤⑥⑦⑧⑨⑩
06	How much are you able to help the most difficult students to pass?	①②③④⑤⑥⑦⑧⑨⑩
07	How much can you influence decisions that are made for this school?	①②③④⑤⑥⑦⑧⑨⑩
08	How much are you able to promote learning when there is little support at home?	①②③④⑤⑥⑦⑧⑨⑩
09	How much can your individual effort influence your students' performance?	①②③④⑤⑥⑦⑧⑨⑩
10	How much can you motivate students who show low interest in schoolwork?	①②③④⑤⑥⑦⑧⑨⑩
11	How much are you able to get students to collaborate on group tasks/projects?	①②③④⑤⑥⑦⑧⑨⑩
12	How much can you inspire confidence and positive attitudes in other educators?	①②③④⑤⑥⑦⑧⑨⑩
13	How much can you inspire hope and confidence in students?	①②③④⑤⑥⑦⑧⑨⑩
14	How much are you able to help students to retain materials that you teach?	①②③④⑤⑥⑦⑧⑨⑩
15	How much can you do to get children to follow class/school rules?	①②③④⑤⑥⑦⑧⑨⑩
16	How much can you freely express your views on important school matters?	①②③④⑤⑥⑦⑧⑨⑩
17	How much are you able to keep your students "on-task" with difficult tasks?	①②③④⑤⑥⑦⑧⑨⑩
18	How much can you do to make your students enjoy coming to school?	①②③④⑤⑥⑦⑧⑨⑩
19	How much are you able to keep your students "focused" during difficult times?	①②③④⑤⑥⑦⑧⑨⑩
20	How much are you able to do to make sure your students learn grade level content?	①②③④⑤⑥⑦⑧⑨⑩

Page 5. *Please go on to the next page.*

This is the final section of the survey.

01	Which class was your primary responsibility (homeroom)? ◯ I did not teach	◯ Preschool	◯ Infant 1	◯ Infant 2
		◯ Standard 1	◯ Standard 2	◯ Standard 3
		◯ Standard 4	◯ Standard 5	◯ Standard 6

02	How many students were there in the class that was your primary responsibility? *(above)*	
03	How many of your students had their basic materials? *(textbooks, notebooks, utensils)*	
04	How much class time was allotted for English daily? *(in minutes)*	
05	How much class time was allotted for Math daily? *(in minutes)*	
06	How many students total were you directly responsible for? (*class, club, competitions, team…*)	
07	FOR THE WEEK How many hours did you spend weekly to write plans (prepare) so you could do your job well?	
08	How many of your students lived in a home with adults who did not understand English well AND the student also struggled with English in their academic work?	
09	Consider the issue of *being poor / poverty* in terms of *"a student who only had one or two uniforms and might have come to school hungry some days because the family was struggling to make ends meet."* Using the description above, how many of those students you were directly responsible for at the school, would *you* have classified as being poor or living in poverty?	

10

2017-2018 GRADES: Please complete this table with your class' average in each of these subjects.

For each term and each subject, add the "report card grade" for students in your class then divide the total by the number of students in your class that term. That will give you the class average for the term in that subject.

Example: English

5 pupils total 1st Term: (Ali 70, Ann 85, Bill 85, Jack 78, Sue 89), [70 + 85 + 85 + 78 + 89] = 407 407 / 5 = **81.40**

4 pupils (Ali moved) 2nd Term: (Ann 77, Bill 88, Jack 80, Sue 85), [77 + 88 + 80 + 85] = 330 330 / 4 = **82.50**

6 pupils (2 new) 3rd Term: (Ann 79 + Bill 86 + Jack 84 + Sue 88 + Ty 90 + Zia 75), [79 + 86 + 84 + 88 + 90 + 75] = 502 502 / 6 = **83.67**

Only put in your class averages in the table below.
For example, the averages for English would be **81.40, 82.50,** and **83.67**

	English	Math	Science	Social Studies
1st term				
2nd term				
3rd term				

Page 6. **End of the survey.**

To help keep your participation status and your responses confidential, please seal the envelope with your survey & raffle ticket, before you return it.

Thank you for investing your time and effort in this process. Your contribution is greatly appreciated.
Remember that as an educator, everything you do matters!
Bo'tic. Danke. Gracias. Seremei. Tangks. Thank you.

Appendix C6: Follow-up Note

Dear Educator,

This is a follow-up request asking if you would be willing to complete survey materials for my dissertation. Your contribution is important for my dissertation study, which has potential for future application in Belize.

Enclosed is another copy of the survey materials. Included in this packet are the survey materials, a sweepstakes ticket, a return envelope, and postage stamps. Please read the informed consent sheet and I would be grateful if you provide a response to all items in the form.

You may return the survey either by postal service or simply drop off the sealed envelope at ***Tech Design*** (#8 Ramos Road) in Dangriga Town. You can be assured that your sealed envelope will be sent to me unopened. Confidentiality is vitally important in this study. I sincerely appreciate your contribution to shaping how we look at and how we can estimate school-level educational achievement in Belize.

Thank you,

Betty Jean Usher-Tate

APPENDIX D

Official District-Level Statistics

School			Primary School		Preschool	
ID	Location	Educators	students	classes	available	students
8A01	1	6	..	..	yes	..
8A02	1	7	137	7	no	0
8A03	1	11	217	9	no	0
8A04	2	12	231	8	yes	29
8A05	2	14	261	11	yes	16
8A06	3	4	..	..	yes	..
8A07	1	14	180	8	yes	38
8A08	2	20	357	17	yes	48
8A09	1	24	291	21	yes	30
8A10	1	22	409	20	yes	0
8A11	2	20	315	16	yes	14
8A12	2	10	..	8	yes	6
8A13	2	11	230	9	yes	17
8A14	2	20	355	16	yes	46
8A15	2	19	154	8	yes	11
8A16	3	16	324	13	yes	33
8A17	3	38	961	33	yes	125
8A18	3	10	105	9	yes	17
8A19	3	20	373	17	yes	30
8A20	3	8	113	6	no	0
8A21	1	8	46	4	yes	23
8A22	1	10	191	8	yes	23
8A23	1	12	..	7	yes	..
8A24	1	32	712	26	no	0
8A25	2	7	64	7	yes	14
8A26	2	11	164	8	yes	36
8A27	2	15	359	10	yes	50
8A28	2	20	524	17	yes	54
8A29	3	15	366	12	yes	34
8A30	3	19	344	16	yes	34
8A31	1	13	211	10	yes	43
8A32	2	13	235	8	yes	21
8A33	2	19	407	16	yes	33
8A34	2	5	71	4	yes	1
8A35	1	1	..	..	yes	..
8A36	1	9	77	4	yes	32
8A37	2	10	191	8	yes	27
8A38	2	5	..	..	yes	..
8A39	2	4	..	..	yes	..

APPENDIX E

Appendix E. Summary of Multilevel Syntax

```
INPUT INSTRUCTIONS
Multilevel Achievement Estimation Protocol (MAEP)
Estimating Achievement with 2-level Multilevel Structural Equation Model
                    ;
DATA: FILE IS [path to text file data set]                              ;
VARIABLE: NAMES ARE [list all variables in text file data set]          ;
          USEVARIABLES ARE CTX EFF OTL PRC SBA RIG SCHOOL               ;
          MISSING ARE                ;
          BETWEEN = RIG              ;
          CLUSTER IS  SCHOOL         ;

ANALYSIS:
          TYPE = TWOLEVEL RANDOM     ;
          ALGORITHM=INTEGRATION      ;
           INTEGRATION = 7           ;
           ESTIMATOR = MLR           ;

MODEL:

Level-1 Within-School
          %WITHIN%
          ACHW BY CTX EFF OTL PRC    ;
           EEB | SBA ON ACHW         ;

Level-2 Between-School
           %BETWEEN%
          ACHB BY CTX EFF OTL PRC    ;
          SBA EEB ON ACHB RIG        ;

OUTPUT: TECH1;
```

APPENDIX F:

Model Estimates

Standardized School-Level Estimates

ID School	School Size (in terms of the number of) Educators	Students	Standardized Achievement Estimates (all schools represented) PSE	BASE	WISP	MAEP	Standardized Achievement Estimates: Listwise deletion (*n*=33) PSE	BASE	WISP	MAEP
1	9	198	0.39	-0.08	-0.04	-1.24	0.43	-0.08	-0.03	-1.25
2	10	231	0.76	1.10	1.05	1.00	0.80	1.11	1.07	1.13
3	12	258	0.23	0.34	0.25	-0.13	0.27	0.35	0.26	-0.07
4	4	31	-0.97	-1.22	-1.28	-0.67	-0.95	-1.22	-1.27	-0.64
5	11	204	-0.24	0.49	0.31	1.28	-0.21	0.49	0.33	1.42
6	10	241	1.11	0.90	0.97	0.40	1.16	0.91	0.99	0.49
7	18	294	1.02	0.97	0.96	0.52	1.06	0.98	0.98	0.61
8	4	72	0.71	0.85	0.80	0.83	0.75	0.86	0.82	0.95
9	38	961	1.09	0.88	1.05	-0.27	1.14	0.89	1.07	-0.22
10	9	193	-0.81	-1.94	-1.70	-2.26	-0.78	-1.94	-1.70	-2.33
11	18	329	0.16	0.58	0.47	0.75	0.20	0.59	0.48	0.86
12	18	339	-0.22	0.26	0.13	0.77	-0.19	0.26	0.14	0.88
13	21	290	-0.26	-0.74	-0.53	-0.63	-0.23	-0.74	-0.52	-0.61
14	22	485	0.65	0.46	0.58	-0.37	0.69	0.47	0.60	-0.32
15	18	363	1.47	0.59	0.82	-0.52	1.52	0.60	0.84	-0.48
16	8	113	-1.52	-0.76	-1.02	0.43	-1.50	-0.76	-1.01	0.52
17	14	339	0.18	-0.44	-0.30	-0.82	0.22	-0.44	-0.29	-0.80
18	8	154	0.80	-0.45	-0.09	-1.38	0.85	-0.45	-0.08	-1.39
19	10	234	0.19	0.16	0.16	0.07	0.23	0.17	0.17	0.14
20			-0.49	-0.97	-0.87					
21	8	137	-0.90	-0.33	-0.51	0.25	-0.88	-0.33	-0.50	0.33
22			1.57	1.16	1.28					
23	8	182	-0.52	0.55	0.27	1.29	-0.49	0.56	0.29	1.43
24	18	352	-1.09	-0.59	-0.70	-0.15	-1.07	-0.59	-0.69	-0.09
25	9	181	1.36	1.53	1.48	1.34	1.41	1.54	1.50	1.49
26	5	58	-2.49	-2.27	-2.50	-1.19	-2.48	-2.28	-2.50	-1.20
27	30	712	0.76	1.75	1.66	1.44	0.80	1.77	1.68	1.59
28	9	163	-0.90	-0.95	-0.91	-0.84	-0.88	-0.95	-0.91	-0.83
29	14	371	-0.74	-1.28	-1.09	-1.32	-0.71	-1.28	-1.08	-1.34
30	11	362	0.03	-0.22	-0.08	-0.55	0.06	-0.21	-0.07	-0.51
31	17	338	-2.24	-2.01	-2.11	-1.33	-2.23	-2.01	-2.11	-1.34
32	19	526	-0.21	0.37	0.23	0.55	-0.18	0.38	0.24	0.65
33	4	45	0.73	0.13	0.30	-0.59	0.77	0.14	0.31	-0.55
34	8	197	-0.61	0.19	0.01	0.54	-0.58	0.19	0.02	0.64
35	8	100	0.98	1.00	0.96	0.72	1.02	1.01	0.97	0.83
36						2.09				

School-Level Achievement estimates: Primary School Examinations (PSE), Base-Model (BASE), Weighted-Indicator Scores Protocol (WISP), and Multilevel Achievement Estimation Protocol (MAEP)