ENGLISH MINORITY STUDENTS IN GHANA: HOW LANGUAGE OF TEST ADMINISTRATION AND REGIONAL DIFFERENCES INFLUENCE READING PERFORMANCE ON THE GHANAIAN ACHIEVEMENT TEST

A Thesis

Presented to

The Faculty of the Department

of Psychology

University of Houston

In Partial Fulfillment

of the Requirements for the Degree of

Master of Arts

By

Stephanie Torres

August 2019

ENGLISH-MINORITY STUDENTS IN GHANA: HOW LANGUAGE OF TEST ADMINISTRATION AND REGIONAL DIFFERENCES INFLUENCE READING PERFORMANCE ON THE GHANAIAN ACHIEVEMENT TEST

Stephanie Torres, B.S.

APPROVED:

Elena L. Grigorenko, Ph.D.
Committee Chair
Department of Psychology
University of Houston

Lesley Hart, Ph.D.
Department of Psychology
University of Houston

Paulina A. Kulesz, Ph.D.
Texas Institute for Measurement, Evaluation, and Statistics
University of Houston

Antonio D. Tillis, Ph.D.
College of Liberal Arts and Social Sciences
Department of Hispanic Studies

ENGLISH MINORITY STUDENTS IN GHANA: HOW LANGUAGE OF TEST ADMINISTRATION AND REGIONAL DIFFERENCES INFLUENCE READING PERFORMANCE ON THE GHANAIAN ACHIEVEMENT TEST

An Abstract of a Thesis

Presented to

The Faculty of the Department

of Psychology

University of Houston

In Partial Fulfillment

of the Requirements for the Degree of

Master of Arts

By

Stephanie Torres

August 2019

ABSTRACT

It is estimated that roughly 40% of the world's population lacks access to education in their native tongues. The impact of language of instruction and NDLL (nondominant language learner) status on academic achievement has received significant attention, but less attention has been given when considering degree of urbanization of where children reside and go to school; thus, careful consideration should be given to how cultural differences across regions in low- and middle- income countries may affect performance patterns when children are being taught in their non-dominant language. Here we consider the impact of language of test administration and degree of urbanization on reading performance on the Ghanaian Achievement Test (GAT) in a large sample (N = 1,309, M_{age} = 9.48 years, SD = 2.26; 45.5% Female) of Ghanaian children. Those included in this study are from a survey project carried out by The Education Quality for All (EQUALL) Complimentary Education Program in 2005. All participants completed demographics measures and assessments measuring different competencies such as nonverbal intelligence, language, and reading ability. Negative binomial and zero-inflated Poisson data models revealed that both language of test administration and degree of urbanization were significantly related to reading subskill ability and noted differences in predicted scores between NDLLs and DLLs (dominant language learners). Effects of the degree of urbanization and language of test administration varied by subskill, contrary to what we hypothesized. We expected a fairly linear relationship between all GAT reading subskill scores and degree of urbanization, but in some cases semi-urban local language speakers outperformed both their urban- and rural-dwelling counterparts.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my committee for their continued support, patience, and enthusiasm with this project. I would also like to thank my committee member, Dr. Paulina A. Kulesz, for her invaluable statistical guidance and for sparing her valuable time whenever I approached her with questions. Similarly, I would like to extend my gratitude to Dr. Lesley Hart, and thank her for her guidance throughout this process. Finally, I would like to thank my committee chair, Dr. Elena L. Grigorenko, for her insightful suggestions, guidance, and belief in my abilities since I first came to the University of Houston. None of this would have been possible without their support

TABLE OF CONTENTS

Introduction

Nondominant Language Learners and Academic Achievement

The prevalence of non-dominant language learners (NDLLs) is increasing worldwide. Some estimate that as much as 40% of the global population does not have access to education in a language with which they are comfortable (Walter & Benson, 2012). In the United States, there were an estimated 4.3 million NDLL students in 2004 to 2005, but this number rose to 4.6 million by 2014, with the number of NDLLs in public schools ranging from 1.0% to 21% depending on the state (Institute of Educational Sciences, 2018). The ever-increasing number of children exposed to multilingual environments in Western countries is paralleled by a surge in research aimed at studying the implications of being an NDLL. One of the objectives of this research is to investigate the association between language proficiency and academic performance. Being an NDLL has repeatedly been shown to be associated with lower reading and math scores when compared to dominant language learners (DLLs) across grade school (Han, 2012). It has also been shown, however, to be associated with secondary educational outcomes equally successful to DLLs when NDLL students are given the opportunity to receive strong, grade-level education in their native language for the first several years of school, through transitional or dual-language classrooms (Collier & Thomas, 2004; Umansky & Reardon, 2014). Being in a dual language immersion classroom has been shown to not only benefit NDLLs, for DLLs have also been noted to have more positive attitudes towards using another language and being more open to interacting with other students unlike themselves even though they do not speak the same language when in dual language classrooms. (Block & Vidaurre, 2019). Unfortunately, instruction in this inclusive fashion is not always available or feasible and children are then placed in fully-immersive classrooms and forced to learn in a language unfamiliar to them. Despite nearly half of the world's population lacking access to education in their native tongues, challenges are most salient in regions where there is a wealth of linguistic diversity, such as in sub-Saharan Africa where reportedly only 13% of its population had access to education in their native tongue in 2000 (UNDP, 2004). As a result, they are at a higher risk of lower long-term academic achievement, especially in the area of reading (Thomas & Collier, 2002).

Academic Achievement/Assessment in Low- and Middle- Income Countries

In 2017, the World Bank warned of a "learning crisis" in global education. Specifically, the organization called for a strong need for policies aimed at developing better learning assessments for use in low- and middle- income countries (LMICs) due to children not being able to demonstrate skills taught in school (World Bank, 2017). It has been suggested that these are due to limitations in assessment materials, which are not able to appropriately capture students' skills and knowledge (e.g., rural inhabitants may be better at applied math vs arithmetic). In both wealthier nations and LMICs, education is complicated, but both researchers and policymakers are attempting to gain a greater understanding on what educational policies will benefit students and their long-term educational outcomes (Glewwe, Hanushek, Humpage, & Ravina, 2011). Educational outcomes across LMICs are highly variable due to a number of factors, such as the quality of education, school resources, and cultural context (Glewwe et al., 2011; Glewwe, Kremer, & Moulin, 2009; Sternberg, 2004). As Sternberg (2004) acknowledged, many programs approach research in a single-culture mindset; thus, it may be the case that the way academic achievement is measured in North

American/European cultures may not apply to students from different ethnic, racial, and national groups (García & Pearson, 1994).

Impact of Degree of Urbanization on Academics

Concerns regarding social inequality arise when one considers the impact urbanization has on education and contributes to disparities in the classroom (Shankar-Brown, 2015). Despite concerns regarding potential educational disparities favoring urbanization, some studies show that contrary to the wide-held belief that rural-residing children are disadvantaged due to environment and limited resources (Barr, 1959), these students may actually benefit from rural living in light of the increased number of social stressors associated with urbanization, such as increased poverty, violence, and marginalization (Logan & Burdick-Will, 2017). Other studies have suggested that children living in more urban environments are at a greater advantage due to a greater number of parents having positive attitudes toward education (e.g., reading) when compared to parents living in rural regions (Memon, Umrani, & Memon, 2016; Ngwaru & Opoku-Amankwa, 2010). Additionally, other studies evaluated children residing in suburban areas (Khattri, Riley, & Kane, 1997) and compared them to children living in rural and urban regions; results suggested that rural and urban children have lower performance scores than children living in suburban areas. For unknown reasons, compared to their urban/central city counterparts, fourth-grade children in rural areas or “small towns” outperform on measures of reading proficiency; when considering those who lived in “urban fringe” and “large towns,” rural-dwelling children outperformed both groups (Campbell, Donahue, Reese, & Phillips, 1996). In conclusion, research examining this achievement gap is inconclusive, and studies have continued to produce mixed results when comparing children who live in rural, semi-urban, and urban regions.

Ghana: Country Characteristics and Its Education System

Ghana is a multilingual country in Africa that has an estimated 81 native languages, of which 73 are indigenous and 8 are non-indigenous (Lewis, 2014). Different languages are spoken across different regions in Ghana, whose economy also differs greatly between those regions (e.g., Northern vs Southern), with the greatest overall differences being noted between those that are urban and rural (Adjei, 2014). These differences are critical because they in turn dictate the educational prospects available to the children of those regions.

Given the diverse linguistic environment of Ghana, special consideration should be given to how language dominance in diverse economic environments can influence reading ability. A recent review of the literature on this topic demonstrated that children in classrooms where the language of instruction (LoI) is a language in which they are not fluent face challenges in learning for a number of reasons, including native language interference and intralingual learning difficulties. These intralingual difficulties include the child making errors that reflect general characteristics of the language, such as the faulty generalization of rules and failing to learn when certain rules apply (e.g., “I before E, except after C”) (Heydari & Bagheri, 2012). These errors are also contingent on the structural differences of the second language, specifically orthographic patterns, phonemic recognition, and direct recognition of words that are already represented in memory (van Staden & Howie, 2012). Nonetheless, in several African countries, including Ghana, the LoI is a European language.

Since English is the LoI in Ghanaian schools, it has been suggested that for children living in communities where the utility of English is minimal (e.g., rural villages), English may be a barrier for learning and a factor that is further limiting their engagement with the curriculum (Milligan & Tikly, 2016; Ngwaru & Opoku-Amankwa, 2010). In addition, it has been reported that local language use is frowned-upon by both peers and teachers, which results in the marginalization of those students who are more fluent in the local language (Hynsjö & Damon, 2016; Maganda, 2016). This oppression is not limited to younger school-aged children though; similar marginalization is reported in higher-education settings among college students and faculty in other countries as well (Zavala, 2014). Overall, it is possible that teachers and other adults reinforce these social hierarchies through language education, but there is potential to challenge these hierarchies and language "regimes," which are shaped by history, political, economic, and other factors (Purkarthofer & De Korne, 2019). This imposition of non-native languages in instructional environments has concerned students, families, and policymakers, who argue that NDLLs will be disadvantaged at school, both at early stages of schooling, while acquiring basic reading skills, and at later stages, while trying to "catch-up" to their dominant-language speaking peers. Educational ministers in Ghana have advocated for the use of a single dominant language as the LoI, with the goal of making the country "globally competitive." Unfortunately, in countries with multiple dialects, these impositions have been a source of grievance associated with a number of sociocultural issues of inequality (UNESCO, 2016).

Degree of Urbanization

Another critical factor to include in the investigation of reading skill development is the degree of urbanization of school location, since the urbanization may contribute to differences in academic achievement. Generally, students residing in urban regions are of higher socioeconomic status, have parents with higher levels of education, and have better access to educational resources and other tangible educational inputs (e.g., computers) than students residing in rural areas (Lounkaew, 2013). Parents shape their expectations of their children based on their prior educational experience. For example, one study reported that fewer rural parents expected their children to acquire a 4-year college degree when compared to parents from suburban or urban regions, even when accounting for differences in socioeconomic status (Lippman, Burns, & McArthur, 1996). As Lounkaew (2013) emphasized, these higher educational expectations correlate with higher levels of achievement. These differences in expectations often dictate the activities in which children are encouraged to engage. For example, children residing in urban areas are often encouraged to engage in educational activities (Memon et al., 2016), while children living in rural communities are often required to do chores and other necessary tasks to help their family and communities (Desai, 2013). In addition, the context in which children learn certain skills while in school may also differ significantly from that of real-world opportunities to practice such skills, depending on where they live and what is expected of them (Tran, 2015). These differences are not limited to children living in LMICs, for the rural-urban achievement gaps have also been noted in high-functioning adults obtaining specialized degrees who live in urban and rural areas (Faisal, Shinwari, & Mateen, 2016); this highlights the importance of considering characteristics of school location and area of residence.

Limitations of Previous Research

As suggested previously, the literature examining the rural-urban achievement gap in developing countries, such as Ghana, is sparse. In addition, it is difficult to make sense of why results are mixed. Most of the literature examining urban and rural differences has been carried out in more higher-income countries (Byun, Meece, & Irvin, 2012; Campbell et al., 1996; Chen, 2012; Lippman et al., 1996; Logan & Burdick-Will, 2017; Mykerezi, Kostandini, & Jordan, 2014; Young, 1998),

Other factors not well studied include socioemotional functioning in students and degree of proficiency in their first language. Previous research suggests that among school-aged children, higher socioemotional functioning is related to higher academic achievement (Niehaus, Adelson, Sejuit, & Zheng, 2017). Niehaus et al. (2017), examined socioemotional functioning across two NDLLs groups (Spanish-speakers and Asian-language speakers) and DLLs (monolingual English speakers) and noted that NDLLs whose dominant language was Spanish reported more socioemotional functioning difficulties than both the Asian-language and English monolingual groups. Additionally, findings examining language proficiency and achievement outcomes of NDLLs have indicated that bilingual children who were most proficient in their native language achieve at higher levels in both their native language and the LoI (Lindholm-Leary & Hernández, 2011). These findings further highlight the importance of considering within-group differences in NDLLs.

Present Study Aims and Hypotheses

The aim of this study is to further extend current knowledge of how children in LMICs perform on measures assessing reading skill when the language of instruction (English) differs from their preferred language (local/native language). We first examined whether dominant-language-learner (DLL; English-dominant) and non-dominant language learner (NDLL; Twi/Ewe/Gonja-dominant) group membership is significantly related to performance on the Ghanaian Achievement Test. Our **first hypothesis** is that language of test administration will be significantly related to reading subtest scores. Additionally, we predict that those children in the DLL group will have higher Ghanaian Achievement Test (GAT) reading scores than all NDLL children, regardless of the local language they speak (e.g., English-dominant child GAT scores > local language-dominant child scores). Given the utility of one language over another depending on area of residence, our **second hypothesis** is that area of residence will also have a significant effect on reading skill performance and have a significant relationship with reading subtest performance. We predict that NDLL children living in rural areas (NDLL-rural) will have the lowest scores of all possible group membership and area of residence combinations (e.g., rural-NDLL, semi-urban-NDLL, urban-NDLL, rural-DLL, semi-urban-DLL, and urban-DLL) across all subtests. We controlled for IQ, expressive vocabulary, age, and schooling status (in- vs out-of-school), given that these may have an effect on our dependent variables of interest and because we are concerned specifically with the effects of our two independent variables on reading performance.

The results of this study can promote discussions about the implications of policies aiming to eliminate the use of native/local languages as the language of instruction in favor of a more “global” language such as English and how such policies relate to children’s reading skill development and performance in the early school years.

Methods

The current study is based off of a larger survey project carried out by The Education Quality for All (EQUALL) Complimentary Education Program funded by the United States Agency for International Development (USAID) in 2005. The larger project’s goal was to develop a foundation for the identification and quantification of children with learning disabilities and other types of special education needs across targeted districts in Ghana.

Researchers from Yale University were engaged to assist with this survey being done in Ghana by USAID.

Ghana is an African country with a current population of 28.2 million. The country is divided into 10 administrative regions (Owusu-Ansah, 2014); three of them (West Gonja, Eastern Region, and Volta) were selected as study sites. More than 80 languages are spoken in Ghana, with English being the only state language; yet, there are 10 government-sponsored languages that represent the majority of the indigenous people of Ghana. Of these languages, Gonja is primarily spoken in Damango; Twi, in the Eastern Region; and Ewe, in Volta.

Damango, the capital of the West Gonja district (population 41,180 in 2010), is largely agricultural with the main economic activities in the district revolving around farming, agro-processing, and trading in foodstuffs. Koforidua (population 183,727 in 2010), in Eastern Ghana, is a largely urban commercial center for the eastern region of the country and serves as a major hub of education in Ghana being home to many schools. Finally, Ho is the capital city of the Volta regions (population 177,281), which, similar to Damango, is largely agricultural with 70% of its economically active labor force being involved in a form of agricultural-related activity (Ghana Statistical Service, 2010). Literacy rates ranged from 52.3% in the region of West Gonja to 90.3% in the Eastern Region, with literacy rates being higher in more urban regions. It was estimated that up to 51.3% of children from rural regions never attended school (Ghana Statistical Service, 2010).

Participants

The sample included 1309 children (415 from Damango, 456 from Koforidua, and 438 from Ho) ages 6 to 15 (M = 9.48, SD = 2.26). Primary-school-aged children were the focus of this study, thus exclusionary criteria included those children that were not between the first and fourth grade. Due to known challenges in recruiting children and families living in isolated regions, they were not actively targeted for recruitment. Both children in (i.e., currently enrolled) and out (currently not enrolled) of school were sampled as part of the larger study. Only larger schools with a sufficient number of children speaking the local languages were targeted for recruitment. To be included in the study, it was necessary for children to speak either English, Ewe, Gonja, or Twi. Across the three study regions, children were recruited primarily from either rural, semi-urban, or urban areas. The EQUALL Complimentary Education Program officer was asked to gather both registered (in-school children) and non-registered out-of-school children from the selected communities. For the purpose of the present investigation, inclusion criteria included completion of the Ghanaian Achievement Test Reading Recognition (N = 1235) and Reading Comprehension (N = 1229) measures.

Procedures (by region)

Damango: Out-of-School Children. Both children and their parents/caregivers were asked to report at the village school at a specified time and date. During the first days of active recruitment, all children 6 to 15 years of age were admitted, while on the last days, girls were favored given their underrepresentation in the sample. Children were asked about language proficiency during the recruitment process and if they spoke neither English or Gonja, the language of the region, they were given a toffee candy and dismissed.

Damango: In-School Children. Recruitment took place at one school in an urban region of Damango. In each classroom, the research staff, with the permission of the school, asked that girls and boys be lined up separately by grade (grades one through four), making sure that they did not line up by height or age, and then selected every *fifth* child until reaching the total number of children required. In Damango, there were two streams per grade, so 25 children were selected, 12 girls and 13 boys, in one stream and 25, 13 girls and 12 boys, in the other. After being selected the children were asked about their language proficiency during the "check-in" procedure, and if a child did not speak either English or Gonja, the child was switched out with another child in the classroom who was randomly selected and was of the same gender.

Damango: Parents. All parents, or guardians/caretakers, of participating children who came to the recruitment event were asked to participate if they spoke either English or Gonja.

Koforidua. Out-of-School Children. Eligible out-of-school children were contacted via a village organizer who arranged to have the children come to a specified location. Initial turnout of children was low as the message did not reach the families of the out-of-school children. A re-announcement was made, which resulted in children trickling in throughout the day. In the following days, children from an average of three villages per day were assessed at a central location. Transportation was made available for parents or caregivers and children when needed. Due to low recruitment of eligible out-of-school children, it became necessary to look for out-of-school children in more remote locations. Since 200 out-of-school children could not be inducted 100 children from a rural school were assessed. This sample of children were separately labeled as rural for the purposes of future analyses.

Koforidua. In-School Children. Children from eight schools participated. For each classroom, the research staff used the same selection procedure described with the Damango sample (see above). In Koforidua, as in Damango there were two streams per grade, so the team selected 25 children (12 girls and 13 boys) in one stream and 25 (13 girls and 12 boys) in the other. Language proficiency was assessed during the "check-in" procedure. All children spoke either English, Twi, or both; therefore, no children needed to be replaced.

Koforidua. Parents. All parents or caregivers of participating children who presented themselves during recruitment were asked to participate, provided they spoke either English or Twi. No parents or caretakers needed to be excluded because of a language incompatibility.

Ho. Out-of-School Children. Prior to testing, members of the research team approached village chiefs and organizers to arrange eligible out-of-school children to meet at a prearranged place for recruitment. Enrollment of out-of-school children and their parents/caregivers was poor, as there are fewer children that do not attend school in Ho than in Damango and Koforidua. Children from an average of five villages per day were assessed at a central location. Transportation was made available for families if needed.

Ho. In-School Children. Four classrooms participated. As done with the aforementioned sites, girls and boys were lined up with boys on one side and girls on the other. Every fifth student was selected from the two lines until reaching the total number of children required. In Ho, there were 3 streams per grade, so the team selected 17 children (8 girls and 9 boys) in one stream, 17 children (9 girls and 8 boys) in the second stream, and 16 children (8 boys and 8 girls) in the last stream to participate. Like in Damango and Koforidua, language was assessed at check-in. All children spoke Ewe, English, or both languages; no children needed to be replaced.

Ho. Parents. All parents or guardians of participating children who came to their child's school during recruitment procedures were also asked to participate in the study. As with the sample in Koforidua and Damango, none of the parents or other caretakers needed to be excluded because of language incompatibilities.

Assessments

Translation of assessment instruments. All assessment materials were translated into three local languages in addition to English: Ewe, Gonja, and Twi. The translations and back-translations were performed by members on the research team who were fluent in both English and at least one of the three languages. Inconsistencies were resolved by the translation and assessment teams and the solutions were homogenized in the three languages.

Assessment Instruments

Demographic information and language dominance. The demographics collected as part of the large epidemiological study refers to the background information that was collected about each participant. Information that was collected includes: age, sex, grade, schooling information (in-school/out-of-school), region of residence, and information on language proficiency/dominance. Language proficiency/dominance was assessed to help determine in which language the assessments should be administered.

The following questions were asked to assess language dominance:

1) What language do you speak at home?

2) What language do you use to ask questions?

3) What language do you use to answer questions?

4) What language do you use to do math?

5) What language do you use to read words?

6) What language do you use to read text?

A total score was calculated to determine language of dominance, which was the language in which the GAT was administered. If the dominant language, or the only language in which the child could read, was not English, Ewe, Gonja, or Twi, the child was excluded from further assessment. If the child was equally fluent in two languages, the administration language was randomly assigned. Once a language was selected, all assessments requiring verbal performance were administered in that language.

The Ghanaian Achievement Test. The Ghanaian Achievement Test (GAT) was derived from achievement assessments previously used by this research team in Zambia (Stemler et al., 2009), constructed to evaluate competencies in four core academic areas: 1) Reading Recognition (letter and word), 2) Pseudo-Word Decoding, 3) Reading Comprehension, and 4) Mathematics. The items were designed to match the concepts expected to be familiar to children in Africa as specified by the government educational standards. All the items were arranged in order of increasing difficulty, matching school grade progression.

The GAT is individually administered and is aimed at quantifying academic skills in children in grades one through seven. To make it accessible to illiterate children, all instructions were read to children by assessors in the language in which the GAT was to be administered. Prior to the GAT administration, assessors ensured that the child understood the expected response type (e.g., that he/she should *point* to the correct answer for each item on the math subtest and on the early reading recognition items). Responses were recorded by assessors onto answer sheets. For those items with four possible multiple-choice answer choices, the student's response choice was recorded, while those items that required a verbal response, were scored on the quality of the response.

The *GAT Reading Recognition* subtest consists of 120 items generating two different types of responses. For the first 60 items examinees are asked to point to 1 of 4 possible answers presented in the test booklet. The first few items begin with letter matching and letter discrimination (*orthography*), and then progress to sound matching, sound discrimination, and then letter-sound correspondence (*phonology*). For the next 60 items, the student was to read aloud single words that progressively became more difficult. Since the skills on the Reading Recognition subtest build on each other, we examined performance on those items assessing orthography (letter matching and letter discrimination), phonology (letter-sound correspondence), and single-word reading.

The *Pseudoword Decoding* subtest is comprised of 38 phonetically regular pseudowords. Initially, the items consist of simple vowel-consonant combinations, such as *ig* or *ak*, and progressively become more difficult in their length and phonetic content.

The *Reading Comprehension* subtest was constructed as a performance response assessment. Students read a word or phrase presented to them on each page of the stimulus booklet and perform the action that the phrase/word instructed. There is a total of 24 items in this subtest with the items becoming progressively more difficult through vocabulary and sentence structure.

Expressive Vocabulary. This assessment is a measure of expressive vocabulary, which refers to the words that the examinee can express or produce by speaking (Burger & Chong, 2011). The participants are shown a series of 47 black and white images and are to name the object presented. There are no time limits. Responses are scored as correct when the child correctly names the object or incorrect if they fail to provide an adequate response.

The Universal Non-Verbal Intelligence Test (UNIT): Cube Design & Symbolic

Memory. The UNIT-1st Edition (Bracken & McCallum, 1998) is a multidimensional nonverbal assessment of intelligence that is used to assess the cognitive abilities of children and adolescents between the ages of 5 to 17 years of age. The UNIT is an assessment that can be administered without using any words, with the goal of ensuring that the child understands what he/she is required to do without giving away the answer; thus, the examiner can use any gestures, symbols, or words that are appropriate to ensure that the child has a clear understanding of what is required of him/her. The UNIT was designed as a "good" cross-cultural assessment that has been adequately studied with individuals from different cultures. For the purpose of this study, a non-verbal intelligence quotient (IQ) composite score was calculated using the Symbolic Memory and Cube Design scores.

The *Symbolic Memory* subtest assesses the child's short-term visual memory and complex sequential memory. It depicts a sequence of universal symbols (baby, girl, boy, man) that is presented to the child for five seconds. After viewing, the child is to re-create the sequence that was presented previously using the *Symbolic Memory* response cards. This assessment is comprised of 30 scored items, 4 demonstration items, and 4 sample items.

The *Cube Design* subtest measures the child's visual-spatial reasoning skills. Each item on this subtest is an abstract, geometric design constructed of green-and-white cubes. Most of the designs presented, with the exception of the first two demonstration items, are three-dimensional. The child is presented with a design that he/she is to recreate using the blocks within a specified time-limit. For each item the examiner recorded the completion time and the total number of points earned for each item.

Quality Assurance and Data Checking

Data quality assurance was performed at the end of each testing day. Two designated team members reviewed all physical copies collected and arranged all assessments by participant number and date of completion, and logged the number of children, parents, and teachers surveyed each day by location.

Data Analytic Plan

All analyses were done using SAS Version 9.4 software (SAS, Inc., Cary, NC) and SPSS version 24 (IBM Corp, 2016). Descriptive statistics were calculated for student characteristics and GAT scores by reading skill (e.g., phonology, pseudoword reading, reading comprehension). 4 of our 6 outcome variables, Phonology, Pseudoword, Word Reading, and Reading Comprehension, were positively skewed with an excess of zeroes and required the need to use a mixed model such as a zero-inflated Poisson (ZIP) model or a zero-inflated negative binomial (ZINB). Using the GENMOD procedure of SAS Version 9.4, we fit the ZIP model to our outcome variables of interest, phonology, pseudoword reading, word reading, and reading comprehension. After setting up our model we saved the predicted values and estimated conditional zero-inflation probabilities, the likelihood a child would score a zero or non-zero score. These were then used to generate graphs to examine and determine whether the model is appropriate. Goodness-of-fit statistics were then generated so that we could carry out a test assessing for overdispersion. If overdispersion was detected then we would repeat the aforementioned steps, but instead fit a ZINB model due to it being able to appropriately account for the high number of zeroes while also providing a more flexible estimator for the variance of the outcome variables of interests. We modeled this variation using least square means to appropriately adjust for the other effects of the model. Effect plots were also generated to provide a graphical representation of the interaction effects. We predicted that language of test administration, degree of urbanization, and their interaction would be significantly related to reading subtests scores. We performed a one-way ANCOVA to determine whether language of test administration (English vs local language) predicted total scores for each reading subskill. Additionally, non-parametric analyses were utilized to examine orthography score differences between our groups of interest. Pairwise comparisons were utilized to examine group differences. Non-parametric analyses were also done to examine differences in the orthography performance given its skewed distribution.

As mentioned previously, nonverbal IQ subtests (Symbolic Memory and Cube Design), expressive vocabulary, age, and schooling status (in- vs out-of-school) as covariates. We controlled for these covariates given that these may have an effect on our dependent variables of interest and because we are concerned specifically with the effects of our two independent variables (language and degree of urbanization) on reading performance. These covariates were selected due to statistically significant associations with reading ability in past work done in children (Johnson, 1973; Johnston, 2010; Majeres, 1999; McNorgan, Alvarez, Bhullar, Gayda, & Booth, 2011; Vlachos, Papadimitriou, & Walla, 2015).

Results

Preliminary analyses revealed an excess of zeroes in four of the six outcome variables: Phonology, Pseudoword, Word Reading, and Reading Comprehension. Therefore, we were unable to carry out our initial planned analyses for our variables of interests. The Orthography and Pre-Reading subskill distributions were bimodal and approximately normal, respectively (Figures 1 & 4). Closer graphical examinations using marginal and conditional distributions of each outcome variable suggests that regardless of which language the test was administered in, there continued to be an excess of zeroes for the Phonology, Pseudoword, Word Reading, and Reading Comprehension, indicating children were not able to read or complete many of these tasks. Employing the graphical examination, differences in distribution were noted for Phonology, when comparing marginal versus conditional distributions. Similarly. Pre-Reading's distribution was noted to differ between marginal and conditional distributions, and the scores of those who were tested in English were bimodally distributed. Graphical examinations of distributions conditional on degree of urbanization also revealed differences in distributions. When examining the distributions contingent on degree of urbanization, differences were noted in modes and skewness.

For our analyses we used a zero-inflated Poisson (ZIP) regression model to accommodate our variables of interest that showed an excess of zeroes (Figures 1- 6). In situations when the ZIP model was not sufficiently flexible to account for the excess number of zeros, we ran zero-inflated negative binomial regression models to determine if it provided a better fit. Using the SAS GENMOD procedure, we fit a zero-inflated Poisson model for those outcome variables that had an excess of zeroes in their distribution: Phonology, Pseudoword Reading (Pseudoword), Word Reading, and Reading Comprehension. In addition to predicting subskill scores, we were interested in predicting the likelihood that a child would score a zero. The models included degree of urbanization, language of test administration, and their interactions. Goodness of fit statistics (e.g., scaled Pearson X^2 criterion) were generated to formally test for overdispersion, which is the presence of greater than expected variability in the data based on the model used (Dean & Lawless, 1989). Testing for overdispersion was done to determine whether the model was appropriate for the intended analyses. This ensured that bias was minimized and allowed us to determine whether it was necessary to use a more flexible function, such as a Poisson mixture or hurdle model (Trivedi, 2014).

Orthography

Preliminary analyses showed the negatively skewed distribution of the orthography outcome variable, for which transformations were unable to correct. Therefore, one-way non-parametric analyses were run to examine whether there is a difference in orthography scores among the groups depending on language of test administration and degree of urbanization. Our outcome variable, which is continuous, was converted into ranks and then averaged across our groups of interest for both language of test administration and degree of urbanization. The distributions of the orthography scores were significantly different depending on the degree of urbanization, using Kruskal-Wallis, $X^2 = 179.56$, $p < .001$, due to the presence of three levels and skewed nature of our data. After applying Bonferroni correction to adjust for multiple comparisons, which resulted in a critical *p*-value of $p < .01$, all our groups' scores were noted to be significantly different from one another: rural vs semi-urban ($U = 136.58$, $p < .001$), rural vs urban ($U = 279.53$, $p < .001$), and semi-urban vs urban ($U = 142.96$, $p < .001$). A one-way ANCOVA was done to analyze differences by language of administration, controlling for our covariates of interest. Our analyses revealed a predicted main effect of language of test administration on Orthography performance ($F(1, 1202) = 5.94$, $p = .02$). In order to compare all group combinations using the Kruskal-Wallis, six unique groups were computed by using both language of test administration (two levels) and degree of urbanization (three levels). After applying Bonferroni correction to adjust for multiple comparisons, which resulted in a critical *p*-value of $p < .008$, the following group comparisons showed significant differences: Urban – English vs Rural – English (p <.001), Urban – English vs Urban – Local Language (p <.001), Rural – Local Language vs Semi-urban – English (p <.001), Rural – Local Language vs Rural – English (p <.001), Rural – Local Language vs Urban – Local Language (p <.001), Semi-urban – English vs Urban – English (p <.001), Semi-urban – Local Language vs Urban – English (p < .001).

Phonology

Preliminary analyses showed that Phonology scores were positively skewed, with a skewness of 1.20 and kurtosis of 0.37 (Figure 2). A total of 422 of our participants had total score of zero on this portion of the reading recognition subtest. Of the 1309 participants, we had complete data on 976 of our participants, which were included in the ZIP regression model. Predictors in both the count and inflation portions of the model were statistically significant (p < .001), thus showing that this model fits the data better than the null model (i.e., the intercept-only model). To assess this, we examined the Pearson chi-square statistic and degrees of freedom to test for overdispersion (X^2 = 1796.40, df = 976). If the model is appropriate and there is no indication of overdispersion (i.e., the variability is consistent with what would be expected for this model) the Pearson X^2 statistic divided by the degrees of freedom is expected to have a value of 1. For this model the observed value is higher than what would be expected (X^2/df = 1.84); therefore, a formal one-sided test for overdispersion was done by computing the probability of observing a larger value of the statistic. The results indicated that there was evidence of overdispersion when using the scaled Pearson X^2 criterion (X^2 = 1796.40, df = 976, p < .001). Nevertheless, when controlling for our covariates, age, sex, nonverbal IQ, expressive vocabulary, and schooling status (in- vs out-of-school), the results as presented indicate that language of test administration ($\beta_{English}$ = 0.57, *SE* = 0.06, Wald 95% CI [0.46, 0.68], Wald X^2 = 105.99, p < .001) is a significant determinant of the expected Phonology value, as is the interaction between language of test administration and degree of urbanization ($\beta_{\text{English x Semi-urban}}$ = -0.62, *SE* = 0.09, Wald 95% CI [-0.80, 0.45], Wald X^2 = 47.51, p < .001). When examining the zero model, or likelihood of whether our covariates of interest are significant predictors of the probability of a child scoring a zero or non-zero value, it was noted that nonverbal IQ (β = -0.02, *SE* = 0.0039 Wald 95% CI [-0.03, -0.01], Wald X^2 = 21.39, p < .001), expressive vocabulary (β = -0.05, *SE* = 0.0095, Wald 95% CI [-0.07, -0.04], Wald X^2 = 32.91, p < .001), and schooling-status (β = 1.00 *SE* = 0.18, Wald 95% CI [0.65, 1.34], Wald X^2 = 32.05, p < .001), were all significant predictors, while sex (p = .28) and age (p = .95) were not significant.

Despite a ZIP model being able to account for the excess number of zeros, the scaled Pearson statistic criterion (X^2 = 1796.40, p <.001) indicated the presence of overdispersion, which suggested that another model needed to be fitted since the ZIP model estimates were unreliable. A zero-inflated negative binomial (ZINB) model was then used given its ability to account for the excess number of zeroes as well as the ZIP model, while also providing a more flexible estimator for the variance of the response variable. The Pearson X^2 statistic (X^2 = 0.86) of the ZINB in this case is much closer to 1 compared to the ZIP model (X^2 = 1.84). The same one-sided formal test, the examination of the scaled Pearson statistic criterion, that was used to assess the ZIP model's goodness of fit was used to assess for overdispersion. In the case of the ZINB, it indicated that we would fail to reject the null hypothesis of no overdispersion (X^2 = 836.24, df = 976, p > .05) indicating that the ZINB model was a better fit for the data than the ZIP. All of the goodness-of fit criteria (X^2/df, Akaike information criteria, or AIC, and Bayesian information criteria, or BIC) favor the ZINB (X^2/df = 0.86, AIC = 4494.94, BIC =4558.58) over the ZIP model (X^2/df = 1.84, AIC = 5255.45, BIC = 5314.20), with smaller values being more desirable for the AIC and BIC criteria. Table 5 provides a side-by-side comparison of the other goodness-of-fit criteria for both the ZIP and ZINB models. The negative binomial dispersion parameter has an estimated value of 0.62 (*SE* = 0.07, Wald 95% CI, [0.50, 0.76]), thus indicating that the estimate is significantly different from 0. This suggests that there is no evidence of overdispersion in the ZINB model; therefore, it is reasonable to assume that the standard errors of the ZINB model's parameter estimates are unbiased and that the model's estimates are suitable for statistical inference. When examining the covariates of interest, it was noted that only nonverbal IQ (β = -0.02, *SE* = 0.01, Wald 95% CI [-0.03, -0.01], p < .001), expressive vocabulary (β = -0.08, *SE* = 0.02, Wald 95% CI [-0.12, -0.05], *p* <.001, and schooling status (β = 1.21, *SE* = 0.24, Wald 95% CI [0.73, 1.68], *p* <.001) were predictive of whether a child would receive a total of a zero or non-zero value on the phonology portion of the GAT. Sex (*p* = .12) and age were not significant (*p* = .75).

Figure 3 shows an effect plot that provides a graphical summary of the relationship between the model's prediction of Phonology scores based on the categorical variables of interest, language of test administration and degree of urbanization. The graph suggests that degree of urbanization has a negative, fairly linear, effect on the expected phonology score regardless of what language the GAT is administered in, with those living in more urban areas having higher scores overall (English: *M* = 7.18, *SD* = 0.80; Local Language: *M* = 3.54, *SD* = 0.42). A modest score difference, was noted in the expected mean phonology scores between urban (*M* = 3.54, *SD* = 0.42) and semi-urban dwellers (*M* = 3.46, *SD* = 1.12) that were local-language dominant, while those living in more developed regions had slightly higher scores. Overall, lower predicted values were noted with those in rural regions regardless of language administration for the GAT. The LS-means were used to assess differences between group means. The interaction between the effects of test language administration and degree of urbanization is significantly clear given the lack of parallelism (Figure 4). The effects of both language of test administration ($F(1, 683) = 10.46$, $p = .001$) and degree of urbanization ($F(2, 683 = 5.64$, $p = .004$) were significant, as was the interaction ($F(2, 683) = 12.28$, $p <.001$).

Reading Recognition – Pre-reading

Preliminary analyses indicated that pre-reading subtest scores were more normally distributed and did not contain a high number of zeroes as did the other outcome variables of interest, although it did have a bimodal shape (Figure 5). Multiple comparisons were done to examine the differences among the mean scores between our groups. Adjusted for all other terms in the model and correcting for multiple comparisons using the Tukey adjustment, the interaction of both degree of urbanization and language of administration were noted to be non-significant ($F(5, 1009) = 1.04$, $p = .35$). Statistically significant differences were found between groups, with regards to both language of test administration ($F(1, 1009) = 46.88$, $p < .001$) and degree of urbanization ($F(2, 1009) = 118.67$, $p <.001$). Post-hoc analyses using Tukey-Kramer adjustment indicated that local language dominant children's scores ($M_{LS} =16.87$) were lower than those whose dominant language was English ($M_{LS} = 20.59$, $p < .001$). Statistically significant differences were noted across several groups (Table 6). Consistent with our hypothesis, English-dominant children living in more urban regions had scores that were significantly different when compared to all the other groups, with their scores being higher overall (Figure 12).

A two-way or 3 x 2 ANCOVA was done to determine whether there is an interaction effect between both degree of urbanization and language of test administration in terms of GAT-Pre-Reading scores, after adjusting for covariates. Results of these analyses indicate that the main effects of both language of test administration ($F(1, 1011) = 187.74$, $p < .001$, $\eta^2=.145$) and degree of urbanization ($F(2, 1011) = 96.37$, $p < .001$, $\eta^2=.074$) were significantly related to Pre-Reading scores when controlling for all other effects. As suggested previously, these analyses also indicate that the overall effect of the interaction of both language of test administration and degree of urbanization when controlling for other effects were not significant ($F(1, 1011) = 160.84$, $p = .11$, $\eta^2=.002$).

Reading Recognition – Word Reading

Preliminary analyses and graphical distributions of Word Reading scores showed a highly positively skewed distribution (Figure 6), with a skewness of 3.24 and kurtosis of 9.83. A total of 988 participants had a total score of 0, thus indicating an excess number of zeros and the need to use a zero-inflated model. Using the PROC GLM SAS procedure we used a ZIP regression and fit it the outcome variable Word Reading. We included the degree of urbanization and language of test administration variables as linear predictors in the model. Goodness of fit statistics were saved in order to carry out a formal test for overdispersion. Overdispersion was noted when using a ZIP regression model ($X^2 = 1592.30$, $df = 978$, $X^2/df = 1.62$); therefore, the estimates are suspect. Nevertheless, using a ZIP regression, the results as presented indicate that both language of test administration ($ß_{English} = 5.31$, Wald $X^2 = 127.10$, Wald 95% CI [4.39, 6.24], $p < .001$), degree of urbanization ($ß_{rural} = 5.40$, Wald $X^2 = 129.77$, Wald 95% CI [4.47, 6.33], $p < .001$; $ß_{semi\text{-}urban} = 4.71$, Wald $X^2 = 98.29$, Wald 95% CI [3.78,5.64], $p < .001$), and their interaction ($ß_{English\ x\ rural} = -5.33$, Wald

X^2 = 123.47, Wald 95% CI [-6.28, -4.39], p < .001; $ß_{English \times semi\text{-}urban}$ = -4.95, Wald X^2 = 107.72, Wald 95% CI [-5.88, -4.01], p < .001) are significant determinants of the Word Reading expected values. Even though a zero-inflated Poisson model can account for the excess of zeroes, the Pearson statistic of the ZIP model suggested model misspecification due to overdispersion. Thus, we fit a zero-inflated negative binomial as a mean to account for both the excess of zeroes by providing a more flexible estimator for the variance of the response variable.

Using the PROC GENMOD procedure, a ZINB model was then fit to the response variable Word Reading. The model specifications were similar to those used in the ZIP regression, including both degree of urbanization and language of test administration variables as predictors in the model. The fit criteria for the ZINB model resulted having a Pearson chi-square statistic (X^2/df = 0.70), more favorable than that of the ZIP Poisson model. A formal follow-up test was done, just as with the ZIP model, and revealed a non-significant result, indicating that the issue of overdispersion was accounted for (X^2 = 684.41, df = 978, p > .05) by using a ZINB Poisson regression model. The negative binomial dispersion parameter has an estimated value of 1.12 (Wald 95% CI [0.86, 1.47]), supporting the use of the ZINB model. Compared to the ZIP model, all the parameter estimates are slightly larger in magnitude and there is no change in any inference overall, thus supporting that language of test administration ($ß_{English}$ = 5.34, Wald X^2 = 111.30, Wald 95% CI [4.35, 6.33], p < .001), degree of urbanization ($ß_{rural}$ = 5.41, Wald X^2 = 86.97, Wald 95% CI [4.27, 6.55], p < .001; $ß_{semi\text{-}urban}$ = 4.67, Wald X^2 = 71.33, Wald 95% CI [3.59, 5.75], p < .001), and their interaction ($ß_{English \times rural}$ = -5.32, Wald X^2 = 52.71, Wald 95% CI [-6.76, -3.89], p < .001; $ß_{English \times semi\text{-}urban}$ = -4.99, Wald X^2 = 69.65, Wald 95% CI [-6.16, -3.82], p < .001) are significant predictors of GAT Word Reading scores, controlling for our covariates of interest. It was noted that nonverbal IQ (β = -0.02, *SE* = 0.004, Wald 95% CI [-0.03, -0.01], p <.001) and expressive vocabulary (β = -0.06, *SE* = 0.01, Wald 95% CI [-0.11, -0.06], p <.001), and schooling status (β = 1.66, *SE* = 0.01, Wald 95% CI [-0.08, -0.04], p <.001) were predictive of whether a child was to earn a total of a zero or non-zero positive value on the word reading portion of the GAT. An effect plot was generated to provide a graphical summary of the relationship between the model's prediction of Word Reading scores. Figure 7 shows an effect plot which provides a graphical summary of the relationship between the model's prediction of Word Reading scores based on the categorical variables of interest, language of test administration and degree of urbanization. It shows that, as expected, degree of urbanization has a negative, approximately linear, effect on the predicted Word Reading score when the language of test administration is English, with those living in urban regions having higher predicted values (*M* =12.19, *SD* = 5.37) than those children living in semi-rural (*M* = 6.57, *SD* = 3.42) and rural regions (*M* = 1.75, *SD* = 0.97). As predicted, children who were administered the GAT in English and lived in more urban regions had the highest predicted values of all groups, followed by those living in semi-urban areas and then rural. On the other hand, when looking at the relationship between local language and degree of urbanization, semi-urban dwellers (*M* = 3.08, *SD* = 2.36) had higher predicted scores than both rural (*M* = 1.29, *SD* =1.21) and urban dwellers (*M* = 0.04, *SD* = 0.02), contrary to what would have been expected. The LS-means were used to assess differences between group Word Reading score means. The interaction between the effects of language of test administration and degree of urbanization is significant, which is suggested given the lack of parallelism in Figure 8. The effect of language of test administration ($F(1, 685) = 21.16$, p <

.001) was significant, as was the interaction between language of administration and degree of urbanization (*F(*2, 685 = 11.63, $p < .001$). Effect of degree of urbanization alone was not significant ($F(2, 685) = 0.15$, $p > .05$).

Pseudoword Reading

Preliminary analyses again showed that the Pseudoword score distribution was also positively skewed (skewness = 2.36; kurtosis: 4.36), with an excess number of zeroes (Figure 9). Therefore, we carried out a ZIP regression. As before, model fit statistics indicated the need to assess for overdispersion ($X^2 = 1968.06$, $df = 978$, $X^2/df = 2.01$). A formal one-sided test was completed and confirmed the presence of overdispersion by referring to the scaled Pearson criterion ($X^2 = 1968.06$, $df = 978$, p < .001). Given the significant criterion, we rejected the null hypothesis of overdispersion not being present; therefore, the estimates were likely to be biased. Despite these potential biases, the ZIP regression results indicated that both language of test administration and degree of urbanization, again, were significant determinants of the expected Pseudoword values, as were two of the interactions, English x Rural ($ß_{\text{English x Rural}} = -0.83$, Wald $X^2 = 32.77$, Wald 95% CI [-1.11, -0.54], $p < .001$) and English x Semi-urban ($ß_{\text{English x Rural}} = -1.35$, Wald $X^2 = 87.55$, Wald 95% CI [-1.63, -1.06], $p < .001$).

A ZINB was then utilized given the presence of overdispersion. The fit criteria were more favorable ($X^2/df = 1.01$) when compared to the ZIP model ($X^2/df = 2.01$). The scaled Pearson X^2 criterion ($X^2 = 985.10$, $df = 978$, $p = .43$) again indicated that the ZINB model is favored over the ZIP model. The dispersion parameter (ß = 1.74) supported this, given that it was both positive and significant (Wald 95% CI [1.27, 2.40]), indicating that the ZINB model fit is more appropriate to use. The ZINB regression results indicated that both language of test administration (ß$_{English}$ = 2.27, Wald X^2 = 65.70, Wald 95% CI [0.28, 1.72], p < .001) and degree of urbanization (ß$_{rural}$ = 1.33, Wald X^2 = 19.49, Wald 95% CI [0.74, 1.92], p < .001; ß$_{semi\text{-}urban}$ = 1.70, Wald X^2 = 25.79, Wald 95% CI [1.04, 2.35], p < .001) were significant determinants of the expected values as were the interactions, while accounting for our covariates. When examining the zero model, or likelihood of our covariates of interest being significant predictors of the probability that a child would score a zero or non-zero value, it was noted that both nonverbal IQ scores (β = -0.02, *SE* = 0.005, Wald 95% CI [-0.03, -0.01], p < .001) and expressive vocabulary (β = -0.04, *SE* = 0.01, Wald 95% CI [-0.06, -0.02], p <.001) were predictive of whether a child was to receive a zero or a positive score. An effect plot was generated manually to graphically summarize the relationship between our model's prediction and categorial predictor. Consistent with our hypothesis, urban dwellers who were administered the GAT in English had higher predicted scores, suggesting that both language and degree of urbanization have an effect on predicted Pseudoword scores. Figure 10 shows that, as expected, degree of urbanization has a negative, roughly linear effect on the expected Pseudoword Reading score when the language of test administration is English and comparing urban to semi-urban scores. Lastly, contrary to our initial hypotheses, English-dominant semi-urban dwellers had a lower predicted Pseudoword score (*M* = 4.74, *SD* = 1.55) than the rural dwellers (*M* = 5.38, *SD* = 1.57). LS-means were used to assess group mean differences. The interaction between the effects of language of test administration and degree of urbanization is significant given the lack of parallelism in the interaction plot (Figure 11). After applying Bonferroni correction to adjust for multiple comparisons, the effects of language of test administration ($F(1, 685) = 57.97$, p < .001) and the interaction between language of test administration and degree of urbanization ($F(2, 685) = 6.90$, p

<.001) were noted to be significant. Again, degree of urbanization was not significant after accounting for multiple comparisons ($F(2, 685) = 3.68$, $p = .03$)

Reading Comprehension

Examinations of the graphical distribution and descriptive statistics (kurtosis = 10.96, skewness: 3.22) showed an excess of zeroes on the reading comprehension subtest as well (Figure 12), with 853 participants having a total score of zero. A ZIP regression was done that included both language of test administration, degree of urbanization, and their interactions in the model of the linear predictor. Degree of urbanization was used as the linear predictor in the logit model that explores the probability of a zero count. The predicted values and the zero-inflation probabilities are used to assess the model's goodness-of-fit. The Pearson statistic suggested the presence of overdispersion ($X^2 = 2002.21$, $df = 975$, $X^2/df = 2.05$), even though a ZIP model was used to account for the excess of zeroes. A formal one-sided test for overdispersion was performed by computing the probability of observing a larger value of the Pearson statistic. Given the results of this test ($X^2 = 2002.21$, $df = 975$, $p < .001$), we rejected the null hypothesis of no overdispersion at the most commonly used confidence intervals. Therefore, parameter estimates are suspect due to the evidence of overdispersion. Nonetheless, the presented results indicate that both language of test administration ($\beta = 3.00$, $SE = 0.29$, Wald 95% CI [-1.48, -0.36], Wald $X^2 = 109.83$, $p < .001$), degree of urbanization ($\beta_{rural} = 2.27$, $SE = 0.29$, Wald 95% CI [1.69, 2.84], Wald $X^2 = 59.48$, $p < .001$; $\beta_{semi\text{-}urban} = 2.23$, $SE = 0.30$, Wald 95% CI [1.64, 2.83], Wald $X^2 = 54.41$, $p < .001$), and their interaction ($\beta_{English\ x\ Rural} = -2.82$, $SE = 0.31$, Wald 95% CI [-3.42, -2.22], Wald $X^2 = 85.20$, $p < .001$; $\beta_{English\ x\ semi\text{-}urban} = -2.87$, $SE = 0.31$, Wald 95% CI [-3.48, -2.27], Wald $X^2 = 86.03$, $p < .001$ are significant determinants of the reading comprehension predicted values.

Due to overdispersion being present in the ZIP model, a ZINB model regression was then utilized. A one-sided follow-up test, using the Pearson statistic, indicated that overdispersion continued to be present, even when using a more flexible estimator ($X^2 = 1157.30$, $df = 975$, $p < .001$). Despite this, the negative binomial dispersion parameter has an estimated value of 1.81, which falls within the Wald 95% CI [1.29, 2.55], supporting the use of the ZINB model over the ZIP. In addition to this, the Pearson X^2 statistic divided by degrees of freedom criterion is much closer to 1 in the ZINB model ($X^2 = 1.19$) when compared to the ZIP ($X^2 = 2.05$) model (see table 5 for additional fit statistics), justifying the use of the negative binomial model for our analyses. When examining the zero model, it was noted that both nonverbal IQ scores ($\beta = -0.02$, $SE = 0.005$, Wald 95% CI [-0.03, -0.01], $p < .001$) and expressive vocabulary ($\beta = -0.04$, $SE = 0.01$, Wald 95% CI [-0.06, -0.02], $p <.001$) continued to be predictive of whether a child was to receive a zero or a positive score, while sex ($p = .16$), age ($p = .59$), and schooling status ($p = .79$) were not significant. An effect plot was generated to graphically summarize the relationship between our model's prediction and categorical predictor. Consistent with our hypothesis, English-speaking urban dwellers had higher predicted scores compared to all other groups (Figure 13). Contrary to our expectations, children who were administered the GAT in the local language and were either semi-urban ($M = 1.06$, $SD = 0.42$) or rural dwellers ($M = 0.83$, $SD = 0.34$) had higher predicted values than urban dwellers ($M = 0.15$, $SD = 0.04$) tested in their native tongue. Nonetheless, we can note from the graphical summary that as urbanization decreased from urban to semi-urban, English test administration had a negative effect on the predicted score, but a modest positive effect when one was living in a rural region. On the other hand, the opposite relationship was found when looking at the effect degree of urbanization had on predicted score values when the GAT was administered in the child's local language – it first increased and then modestly decreased. Overall, we noted that the effect of degree of urbanization on the expected reading comprehension scores was complex across both languages. LS-means were used to assess group mean differences. The interaction between the effects of language of test administration and degree of urbanization was significant given the lack of parallelism, as shown in Figure 14. After applying Bonferroni correction to adjust for multiple comparisons and using $p < .008$ as our critical p-value, the fixed effects of language of test administration ($F(1, 682) = 61.72$, $p < .001$), degree of urbanization ($F(2, 682 = 5.36$, $p = .005$), and their interaction ($F(2, 682) = 6.17$, $p = .002$) were significant.

Discussion

The present investigation examined the roles of language of test administration, degree of urbanization, and their interaction in relation to reading skills in a large sample of children across three districts of Ghana. We hypothesized that those children who received the GAT in

English (DLLs) would have higher scores than those that received the GAT in their local language (NDLLs), and this was confirmed, for language of test administration was shown to have a significant effect on basic reading skills. Across most subskills, it was noted that degree of urbanization also had a significant effect on basic reading performance. Lastly, it was noted that the interaction between degree of urbanization and language of test administration also had a significant role in relation to reading scores, although this effect was more complex and varied by reading skill of interest (e.g., Phonology, Word Reading,

Reading Comprehension). The complex effects of the interaction indicate that there are other factors contributing to our results, which may in fact support greater rural student performance on certain reading subskills (e.g., pseudoword reading) but not others (e.g., phonology).

The results of the present study should be interpreted in light of several limitations. As mentioned previously, the data were collected as part of a survey study by USAID across a subset of the sample in Ghana. Ghana is rich in culture and this diversity in culture can also be seen within different regions of the same country, for there is also a wide range of economic diversity across these regions, with those living in rural regions having more livelihood instability (Dzanku, 2015). Differences in livelihood could be driving some of differences in performance between different areas despite being tested in the child's dominant language. It also could be that, as the World Bank (2017) had mentioned, assessments developed to assess academic achievement in LMICs may not be capturing these children's knowledge or intelligence. Future studies can benefit by collecting qualitative data on how children in rural and urban regions apply the knowledge they learn at school to their day-to-day lives and how it relates to their overall adaptive functioning as defined by people of their town and culture; thus, taking an emic (i.e., within a culture) rather than etic (i.e., between cultures) approach. Other studies have found that both differences in child rearing and development in children from neighboring European countries resulted in varying levels of child adaptive behaviors (Taverna, Bornstein, Putnick, & Axia, 2011), thus supporting the need to take cultural experience into account. Future work can then use themes generated from qualitative data collection to inform assessment development.

In addition to this, future survey studies can benefit from collecting data on the quality of the education across regions in schools, for the present study could have benefited from this information. It may be school and education quality is related to the degree of urbanization, which in turn affects children's learning, which could help explain why rural Ghanaian children had lower scores overall. The goal of USAID's survey was to be as inclusive as possible, but study recruitment only included 3 of 16 regions. Given Ghana's diversity, other studies done in similar countries could benefit from collecting smaller samples from a greater number of regions with the support of local organizations to assist with recruitment efforts. Anecdotally, it was also reported by team members that most of the children who participated in the study were higher functioning. It is believed that due to stigma, locals with children who were atypically developing did not reach out to the research group. Thus, the current data may not fully represent and capture the true academic and developmental diversity in these regions. To address this limitation, future studies could benefit from utilizing a community-based participatory research (CBPR) approach when trying to work with a population that is vulnerable and not much is known about. For known strengths of CBPR include: (1) allowing for adaptations of current resources available in the community, (2) allows researchers to explore knowledge and perceptions held in the population of interest, (3) empowers the community by acknowledging them as the experts on their community, (4) allows for collaborations with community leaders on studying areas of concern/interest in their communities, and (5) helps build trust between the community and researchers (Holkup, Tripp-Reimer, Salois, & Weinert, 2004).

With these limitations in mind, the present investigation expanded on the literature in several ways. First, the majority of studies examining rural and urban differences have been in more economically developed countries (Chen, 2012; Faisal et al., 2016; Mykerezi et al., 2014; Yeung, Craven, & Ali, 2013) and some developing (Tayyaba, 2012), but the research is limited when it comes to reading subskill development in children living in Africa. Most of the current literature available is limited when it comes to literacy in Africa, for the majority of research has focused on medical/sexual health literacy. This study thus helps fill that void by specifically investigating the effect of degree of urbanization and language of test administration on GAT reading subskill scores in a large sample of Ghanaian school-age children. In addition to this, most studies to date have focused on academic achievement differences in rural versus urban dwellers *or* NDLLs (children learning in their second language) versus DLLs (children learning in their first language), but not the combination of both. The current study aimed to fill that gap in the literature as well, with the goal of also highlighting that there are more things to consider when assessing academic achievement in a LMICs, such as culture.

Conclusions

Findings from the present study revealed a meaningful, albeit complex, main effect of both language of test administration, degree of urbanization, and their interaction on reading subskills among a large sample of Ghanaian school-age children. Overall, children in rural regions were noted to have lower scores across all subtest scores, suggesting that they are not able to show what they are learning, if they are learning at all. When examining the predicted value scores of the interaction between degree of urbanization and language of test administration, this changed, for in some cases rural-dwelling children outperformed their urban- and semi-urban dwelling counterparts, which is contrary to what was expected. Additional research is warranted to examine how children who live in rural regions apply the knowledge they acquire at school and how cultural differences across different regions may support these children.

In conclusion, when looking at NDLL-status, these children had lower scores overall, but environmental differences, in some cases, appeared to support these children who do not live in urban areas; therefore, it is suggested that NDLL-status does not necessarily “doom” a student to fail, for environmental factors can help mitigate risks. Children that were dominant in the LoI, English, all had higher scores when compared to those who were local-language dominant, indicating that they may be benefiting more from their schooling. This finding alone indicates that there are classroom disparities due to language; thus, it is important to consider student demographic factors, such as language, since they are central in supporting student academic success (Pollock, 2008). The imposition of “language regimes” do not allow children to learn and practice in their native language, thus leaves them feeling unsupported and marginalized, as both Hynsjö (2016) and Purkarthofer (2019) explained. To address classroom disparities, it has been shown that embracing a children’s cultural identity and allowing them to use their home language/native tongue promotes thoughtful, inclusive interactions, which in turn enhances NDLL engagement with material (Kumar, Zusho, & Bondie, 2018); thus, it is recommended that educators and policymakers strive to understand how students’ backgrounds influence their academic performance and consider this when making decisions about curriculum. Ultimately, it is policymakers’ responsibility to be mindful and work towards understanding how the educational and societal structures they establish may inadvertently disadvantage some students and privilege others.

References

Adjei, E. K. M., S. O. (2014). Regional Income Differences in Ghana: The Importance of Socio-Demographic and Ethnicity. *Research on Humanities and Social Sciences, 4*(17).

Asante, K.B. (2017, January). Keep English language as medium of instruction in schools. *Ghana Web*. Retrieved from https://www.ghanaweb.com/GhanaHomePage/News Archive/Keep-English-language-as-medium-of-instruction-in-schools-K-B-Asante-503892

Barr, F. (1959). Urban & Rural Differences in Ability & Attainment. *Educational Research, 1*(2), 49-60. doi:10.1080/0013188590010204

Burger, A., & Chong, I. (2011). Receptive Vocabulary. In S. Goldstein & J. A. Naglieri (Eds.), *Encyclopedia of Child Behavior and Development* (pp. 1231-1231). Boston, MA: Springer US.

Block, N., & Vidaurre, L. (2019). Comparing attitudes of first-grade dual language immersion versus mainstream English students. *Bilingual Research Journal, 42*(2), 129-149. doi:10.1080/15235882.2019.1604452

Byun, S. Y., Meece, J. L., & Irvin, M. J. (2012). Rural-Nonrural Disparities in Postsecondary Educational Attainment Revisited. *Am Educ Res J, 49*(3). doi:10.3102/0002831211416344

Campbell, J. R., Donahue, P. L., Reese, C., & Phillips, G. (1996). *NAEP 1994 Reading report card for the nation and the states*. Washington, DC: National Center for Education Statistics

Chen, Y.H. (2012). Cognitive diagnosis of mathematics performance between rural and urban students in Taiwan. *Assessment in Education: Principles, Policy & Practice, 19*(2), 193-209. doi:10.1080/0969594x.2011.560562

Collier, V. P., & Thomas, W. P. (2004). The astounding effectiveness of dual language education for all. *NABE Journal of Research and Practice, 2*, 1-20.

Dean, C., & Lawless, J. F. (1989). Tests for Detecting Overdispersion in Poisson Regression Models. *Journal of the American Statistical Association, 84*(406), 467-472. doi:10.1080/01621459.1989.10478792

Desai, Z. (2013). Local languages: Good for the informal marketplace but not for the formal classroom? *Education as Change, 17*(2), 193-207. doi:10.1080/16823206.2013.803659

Dzanku, F. M. (2015). Transient rural livelihoods and poverty in Ghana. *Journal of Rural Studies, 40,* 102-110. doi:10.1016/j.jrurstud.2015.06.009

Faisal, R., Shinwari, L., & Mateen, H. (2016). Evaluation of the academic achievement of rural versus urban undergraduate medical students in pharmacology examinations. *Asian Pacific Journal of Reproduction, 5*(4), 317-320. doi:10.1016/j.apjr.2016.06.002

García, G. E., & Pearson, P. D. (1994). Assessment and Diversity. *Review of Research in Education, 20,* 337-391.

Ghana Statistical Service. (2014). *2010 Population & Housing Census: Eastern Region.*

Ghana Statistical Service. (2014). *2010 Population & Housing Census: Ho Municipality.*

Ghana Statistical Service. (2014). *2010 Population & Housing Census: West Gonja District.*

Glewwe, P. W., Hanushek, E. A., Humpage, S. D., & Ravina, R. (2011). School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010. *Education Policy in Developing Countries*, 13-64. doi:10.3386/w17554

Glewwe, P. W., Kremer, M., & Moulin, S. (2009). Many Children Left Behind? Textbooks and Test Scores in Kenya. *American Economic Journal: Applied Economics, 1*(1), 112-135. doi:10.1257/app.1.1.112

Han, W. J. (2012). Bilingualism and academic achievement. *Child Dev, 83*(1), 300-321. doi:10.1111/j.1467-8624.2011.01686.x

Heydari, P., & Bagheri, M. S. (2012). Error Analysis: Sources of L2 Learners' Errors. *Theory and Practice in Language Studies, 2*(8). doi:10.4304/tpls.2.8.1583-1589

Holkup, P. A., Tripp-Reimer, T., Salois, E. M., & Weinert, C. (2004). Community-based participatory research: an approach to intervention research with a Native American community. *ANS. Advances in nursing science, 27*(3), 162-175. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/15455579

Hynsjö, D., & Damon, A. (2016). Bilingual education in Peru: Evidence on how Quechua-medium education affects indigenous children's academic achievement. *Economics of Education Review, 53*, 116-132. doi:10.1016/j.econedurev.2016.05.006

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

Institute of Educational Sciences. (2018). *English Language Learners in Public Schools*. Retrieved from: https://nces.ed.gov/programs/coe/indicator_cgf.asp

Johnson, D. D. (1973). Sex Differences in Reading across Cultures. *Reading Research Quarterly, 9*(1), 67 - 86.

Johnston, R. (2010). Investigating gender differences in reading AU - Logan, Sarah. *Educational Review, 62*(2), 175-187. doi:10.1080/00131911003637006

Khattri, N., Riley, K. W., & Kane, M. B. (1997). Students At Risk in Poor, Rural Areas: A Review of the Research. *Journal of Research in Rural Education, 13*(2), 79-100.

Kumar, R., Zusho, A., & Bondie, R. (2018). Weaving Cultural Relevance and Achievement Motivation Into Inclusive Classroom Cultures. *Educational Psychologist, 53*(2), 78-96. doi:10.1080/00461520.2018.1432361

Lewis, M.P., Simons, G.F., Fennig, C.D. (2014). Ghana in *Ethnologue: Languages of the World*. 17th ed. Murica, TX: SIL International.

Lindholm-Leary, K., & Hernández, A. (2011). Achievement and language proficiency of Latino students in dual language programmes: native English speakers, fluent English/previous ELLs, and current ELLs. *Journal of Multilingual and Multicultural Development, 32*(6), 531-545. doi:10.1080/01434632.2011.611596

Lippman, L., Burns, S., & McArthur, E. (1996). *Urban Schools: The Challenge of Location and Poverty*. Retrieved from Washington, DC:

Logan, J. R., & Burdick-Will, J. (2017). School Segregation and Disparities in Urban, Suburban, and Rural Areas. *Ann Am Acad Pol Soc Sci, 674*(1), 199-216. doi:10.1177/0002716217733936

Lounkaew, K. (2013). Explaining urban–rural differences in educational achievement in Thailand: Evidence from PISA literacy data. *Economics of Education Review, 37*, 213-225. doi:10.1016/j.econedurev.2013.09.003

Maganda, D. (2016). Conflicted Worlds of Multilingual Communities in Africa: Literacy Tangled in Words. *Journal of Language and Literacy Education, 12*(1), 53-78.

Majeres, R. (1999). Sex differences in phonological processes: Speeded matching and word reading. *Memory & Cognition, 27*(2), 246-253. doi:10.3758/BF03211409

McNorgan, C., Alvarez, A., Bhullar, A., Gayda, J., & Booth, J. R. (2011). Prediction of reading skill several years later depends on age and brain region: implications for developmental models of reading. *J Neurosci, 31*(26), 9641-9648. doi:10.1523/JNEUROSCI.0334-11.2011

Memon, S., Umrani, S., & Memon, N. (2016). 'Keep an Eye on Your Son! He Is Reading Too Much These Days': Understanding the Reading Attitudes in English of Rural and Urban Sindhi Students. *Theory and Practice in Language Studies, 6*(8). doi:10.17507/tpls.0608.04

Milligan, L. O., & Tikly, L. (2016). English as a medium of instruction in postcolonial contexts: moving the debate forward. *Comparative Education, 52*(3), 277-280. doi:10.1080/03050068.2016.1185251

Mykerezi, E., Kostandini, G., & Jordan, J. L. (2014). On Rural-Urban Differences in Human Capital Formation: Finding the 'Bottlenecks' *Journal of Rural Social Sciences, 29*(1), 17-47.

Niehaus, K., Adelson, J. L., Sejuit, A., & Zheng, J. (2017). Native language background and academic achievement: Is socioemotional well-being a mediator? *Applied Developmental Science, 21*(4), 251-265. doi:10.1080/10888691.2016.1203790

Ngwaru, J., & Opoku-Amankwa, K. (2010). Home and school literacy practices in Africa- listening to inner voices. *Language and Education, 24*(4), 295-307.

Owusu-Ansah, D. (2014). *Historical Dictionary of Ghana* (4th ed.): Rowman & Littlefield Publishers.

Pammer, K. & Alison, K. (2007). The Contribution of Visual Sensitivity, Phonological Processing, and Nonverbal IQ to Children's Reading AU - *Scientific Studies of Reading, 11*(1), 33-53. doi:10.1080/10888430709336633

Pollock, M. (2008). From Shallow to Deep: Toward a Thorough Cultural Analysis of School Achievement Patterns. *Anthropology & Education Quarterly, 39*(4), 369-380. doi:10.1111/j.1548-1492.2008.00028.x

Purkarthofer, J., & De Korne, H. (2019). Learning language regimes: Children's representations of minority language education. *Journal of Sociolinguistics*. doi:10.1111/josl.12346

Rydland, V., Aukrust, V. G., & Fulland, H. (2012). How word decoding, vocabulary and prior topic knowledge predict reading comprehension. A study of language-minority students in Norwegian fifth grade classrooms. *Read Writ, 25*(2), 465-482. doi:10.1007/s11145-010-9279-2

SAS Institute Inc. (2013). *SAS/STAT Software, Version 9.4*. Cary, NC. URL: http://www.sas.com/.

Shankar-Brown, R. (2015). Urbanization and Persistent Educational Inequalities: The Need for Collective Action Towards Equity and Social Justice. *National Youth-At-Risk Journal, 1*(1). doi:10.20429/nyarj.2015.010104

Stemler, S. E., Chamvu, F., Chart, H., Jarvin, L., Jere, J., Hart, L., . . . Grigorenko, E. L. (2009). Assessing competencies in reading and mathematics in Zambian children. In *Multicultural psychoeducational assessment.* (pp. 157-185). New York, NY, US: Springer Publishing Co.

Sternberg, R. J. (2004). Culture and Intelligence. *American Psychologist, 59*(5), 325-338. doi:10.1037/0003-066x.59.5.325

Taverna, L., Bornstein, M. H., Putnick, D. L., & Axia, G. (2011). Adaptive Behaviors in Young Children: A Unique Cultural Comparison in Italy. *J Cross Cult Psychol, 42*(3), 445-465. doi:10.1177/0022022110362748

Tayyaba, S. (2012). Rural-Urban Gaps in Academic Achievement, Schooling Conditions, Student, and Teachers' Characteristics in Pakistan. *International Journal of Educational Management*, *26*(1), 6-26. doi:10.1108/09513541211194356

Thomas, W. & Collier, V. (2002). *A national study of school effectiveness for language long-term academic achievement.* (CREDE Report). Retrieved from the Center Research on Education, Diversity, & Excellence website: https://escholarship.org/uc/item/65j213pt

Tran, N., Behseta, S., Ellis, M., Martinez-Cruz, A., & Contreras, J. (2015). The Effects of Spanish English Dual Language Immersion on Student Achievement in Science and Mathematics. *eJournal of Education Policy*, 57-77.

Trivedi, P. K. (2014). Models for Count Data. In *Encyclopedia of Health Economics* (pp. 306-311).

Umansky, I. M., & Reardon, S. F. (2014). Reclassification Patterns Among Latino English Learner Students in Bilingual, Dual Immersion, and English Immersion Classrooms. *American Educational Research Journal, 51*(5), 879-912. doi:10.3102/0002831214545110

UNDP. (2004). *Human Development Report 2004: Cultural Liberty in Today's Diverse World*. Retrieved from http://hdr.undp.org/sites/default/files/reports/265/hdr_2004_complete.pdf

UNESCO. (2016). *If you don't understand, how can you learn?* Retrieved from Global Education Monitoring Report: https://unesdoc.unesco.org/ark:/48223/pf0000243713

van Staden, S., & Howie, S. (2012). Reading between the lines: contributing factors that affect Grade 5 student reading performance as measured across South Africa's 11 languages. *Educational Research and Evaluation, 18*(1), 85-98. doi:10.1080/13803611.2011.641270

Vlachos, F., Papadimitriou, A., & Walla, P. (2015). Effect of age and gender on children's reading performance: The possible neural underpinnings. *Cogent Psychology, 2*(1). doi:10.1080/23311908.2015.1045224

Walter, S. L., & Benson, C. (2012). Language policy and medium of instruction in formal education. In B. Spolsky (Ed.), *The Cambridge Handbook of Language Policy* (pp. 278-300). Cambridge: Cambridge University Press.

World Bank (2017). *World Development Report 2018: Learning to Realize Education's Promise*:

The World Bank.

Yeung, A. S., Craven, R. G., & Ali, J. (2013). Self-concepts and educational outcomes of Indigenous Australian students in urban and rural school settings. *School Psychology International, 34*(4), 405-427. doi:10.1177/0143034312446890

Young, D. J. (1998). Rural and Urban Differences in Student Achievement in Science and Mathematics: A Multilevel Analysis. *School Effectiveness and School Improvement*, *9*(4), 386-418. doi:10.1080/0924345980090403

Zavala, V. (2014). What is Quechua Literacy for? Ideological Dilemmas in Intercultural Bilingual Education in the Peruvian Andes. In M. Prinsloo & C. Stroud (Eds.), *Educating for Language and Literacy Diversity: Mobile Selves* (pp. 45-72). London: Palgrave Macmillan UK.

Zero-Inflated Poisson Regression. UCLA: Statistical Consulting Group. Retrieved from https://stats.idre.ucla.edu/sas/output/zero-inflated-poisson-regression/

Table 1.
Participant Characteristics

Number of participants		1309
Mean Age in Years (Range)		9.48 ± 2.26 (6-15)
	(%)	
Sex		
Male	51.9	
Female	48.1	
Schooling		
In-School	66.8	
Out of School	33.2	
Urbanization		
Rural	39.7	
Semi-Urban	23.2	
Urban	37.1	
Language		
English	34.1	
Local Language	60.3	

Table 2.
Correlations for covariates

	1	2	3	4	5
1. Age	-	-.09**	.20**	-0.04	.16**
2. Sex	-.09**	-	-.03	-.14**	-.11**
3. Schooling Status	.20**	-.03	-	-.48**	-.34**
4. Expressive Vocabulary	-.04	-.14**	-.48**	-	.39**
5. Nonverbal IQ	.16**	-.11**	-.34**	.39**	-

Note. $^{**}p<.01$

Table 3.
Scores by Degree of Urbanization

	Subtest	*M* (Range)	*SD*
Urban	Orthography Total Score	16.89 (0-20)	4.40
	Phonology Total Score	6.14 (0-18)	5.44
	Pseudoword Reading Total Score	6.05 (0-38)	11.47
	Pre-Reading Total	23.03 (0-38)	8.14
	Word Reading Total	9.27 (0-76)	18.85
	Reading Recognition Total Score*	32.30 (0-113)	24.01
	Reading Comprehension Total Score	2.80 (0-24)	5.74
Semi-Urban	Orthography Total Score	15.23 (0-20)	4.80
	Phonology Total Score	3.53 (0-17)	4.33
	Pseudoword Reading Total Score	4.79 (0-35)	8.74
	Pre-Reading Total	18.76 (0-36)	7.70
	Word Reading Total	3.68 (0-62)	9.73
	Reading Recognition Total Score*	22.44 (0-96)	15.43
	Reading Comprehension Total Score	1.78 (0-15)	2.95
Rural	Orthography Total Score	12.38 (0-20)	6.19
	Phonology Total Score	1.94 (0-18)	3.31
	Pseudoword Reading Total Score	2.64 (0-38)	6.87
	Pre-Reading Total	14.29 (0-37)	8.19
	Word Reading Total	1.58 (0-73)	8.75
	Reading Recognition Total Score*	15.87 (0-103)	13.76
	Reading Comprehension Total Score	1.10 (0-21)	2.94

*GAT-RR Total Score: The Reading Recognition subtest is comprised of 120 items. The total score (GAT-RR Total) is composed of the orthography (GAT-Orthography), phonology (GAT-Phonology), pre-reading (GAT-RR (pre-reading)), and word-reading (GAT-RR (word reading)) scores listed above.

Table 4.
GAT Scores by Degree of Urbanization and Language of Test Administration

	Subtest	*M* (Range)	*SD*
Urban-LL	Orthography Total Score	16.17 (0-20)	4.65
	Phonology Total Score	3.52 (0-13)	3.53
	Pseudoword Reading Total Score	0.59 (0-30)	3.41
	Pre-Reading Total Score	19.69 (0-33)	6.65
	Word Reading Total Score	0.4 (0-2)	0.23
	Reading Recognition Total Score*	19.73 (0-35)	6.71
	Reading Comprehension Total Score	0.13 (0-10)	1.00
Urban-Eng	Orthography Total Score	17.21 (0-20)	4.25
	Phonology Total Score	7.33 (0-18)	5.73
	Pseudoword Reading Total Score	8.53 (0-38)	12.90
	Pre-Reading Total Score	24.54 (0-38)	8.31
	Word Reading Total Score	13.46 (0-76)	21.47
	Reading Recognition Total Score*	38.00 (0-113)	26.71
	Reading Comprehension Total Score	4.02 (0-24)	6.55
Semi-Urban-LL	Orthography Total Score	13.74 (0-20)	7.74
	Phonology Total Score	3.12 (0-17)	4.36
	Pseudoword Reading Total Score	3.46 (0-34)	7.79
	Pre-Reading Total Score	16.90 (0-36)	8.78
	Word Reading Total Score	2.71 (0-37)	7.63
	Reading Recognition Total Score*	19.61 (0-68)	14.49
	Reading Comprehension Total Score	1.03 (0-15)	2.51
Semi-Urban-Eng	Orthography Total Score	16.50 (0-20)	3.36
	Phonology Total Score	3.89 (0-15)	4.29
	Pseudoword Reading Total Score	5.93 (0-35)	9.35
	Pre-Reading Total Score	20.36 (0-34)	6.24
	Word Reading Total Score	4.52 (0-62)	11.19
	Reading Recognition Total Score*	24.88 (0-96)	15.85
	Reading Comprehension Total Score	2.43 (0-12)	3.16
Rural-LL	Orthography Total Score	12.23 (0-20)	6.22
	Phonology Total Score	1.82 (0-16)	3.06
	Pseudoword Reading Total Score	1.70 (0-38)	5.39
	Pre-Reading Total Score	14.02 (0-37)	8.08
	Word Reading Total Score	1.15 (0-73)	7.65
	Reading Recognition Total Score*	15.16 (0-103)	12.51
	Reading Comprehension Total Score	0.76 (0-21)	2.48
Rural-Eng	Orthography Total Score	13.90 (0-20)	5.48
	Phonology Total Score	2.95 (0-18)	4.88
	Pseudoword Reading Total Score	10.41 (0-35)	11.47
	Pre-Reading Total Score	16.85 (0-37)	8.56
	Word Reading Total Score	5.15 (0-64)	14.69
	Reading Recognition Total Score*	22.00 (0-96)	20.62
	Reading Comprehension Total Score	3.93 (0-19)	4.52

Note. LL = Local language; Eng = English.

*GAT-RR Total Score: The Reading Recognition subtest is comprised of 120 items. The total score (GAT-RR Total) is composed of the orthography (GAT-Orthography), phonology (GAT-Phonology), pre-reading (GAT-RR (pre-reading)), and word-reading (GAT-RR (word reading)) scores listed above.

Table 5.
Comparison of ZIP and ZINB Model Fit Criteria for Outcome Variables

Outcome	**Criterion**	**ZIP**	**ZINB**
Phonology	Pearson Chi-Square	1796.40	836.24
	Pearson X^2 / *df*	1.84	0.86
	Full Log Likelihood	-2615.73	-2234.47
	AIC	5255.45	4494.94
	AICC	5255.78	4495.31
	BIC	5314.20	4558.58
Pseudoword			
	Pearson Chi-Square	1968.06	985.10
	Pearson X^2 / *df*	2.01	1.01
	Full Log Likelihood	-2990.91	-1769.21
	AIC	6005.82	3564.43
	AICC	6006.14	3564.80
	BIC	6064.59	3628.10
Reading			
	Pearson Chi-Square	2002.21	1157.30
	Pearson X^2 / *df*	2.05	1.19
	Full Log Likelihood	-1821.79	-1411.87
	AIC	3667.57	2849.74
	AICC	3667.89	2850.11
	BIC	3726.31	2913.37
Word Reading			
	Pearson Chi-Square	1592.30	684.41
	Pearson X^2 / *df*	1.63	0.70
	Full Log Likelihood	11231.51	-1221.39
	AIC	5559.57	2468.78
	AICC	5559.89	2469.15
	BIC	5618.34	2532.45

Note: Lower AIC, AICC, and BIC values are more desirable.

Table 6.
Least Square Means Pre-Reading Scores for Degree of Urbanization and Language Interaction

	1	2	3	4	5	6
1. Rural – English	-	0.25	0.13	1.00	<.001*	0.34
2. Rural - LL	0.25	-	<.001*	.01	<.001*	<.001*
3. Semi-urban – English	0.13	<.001*	-	.01	<.001*	.98
4. Semi-urban – LL	1.00	0.01	0.01	-	<.001*	.07
5. Urban – English	<.001*	<.001*	<.001	<.001*	-	<.001*
6. Urban – LL	0.34	<.001	.98	.07	<.001*	-

Figures

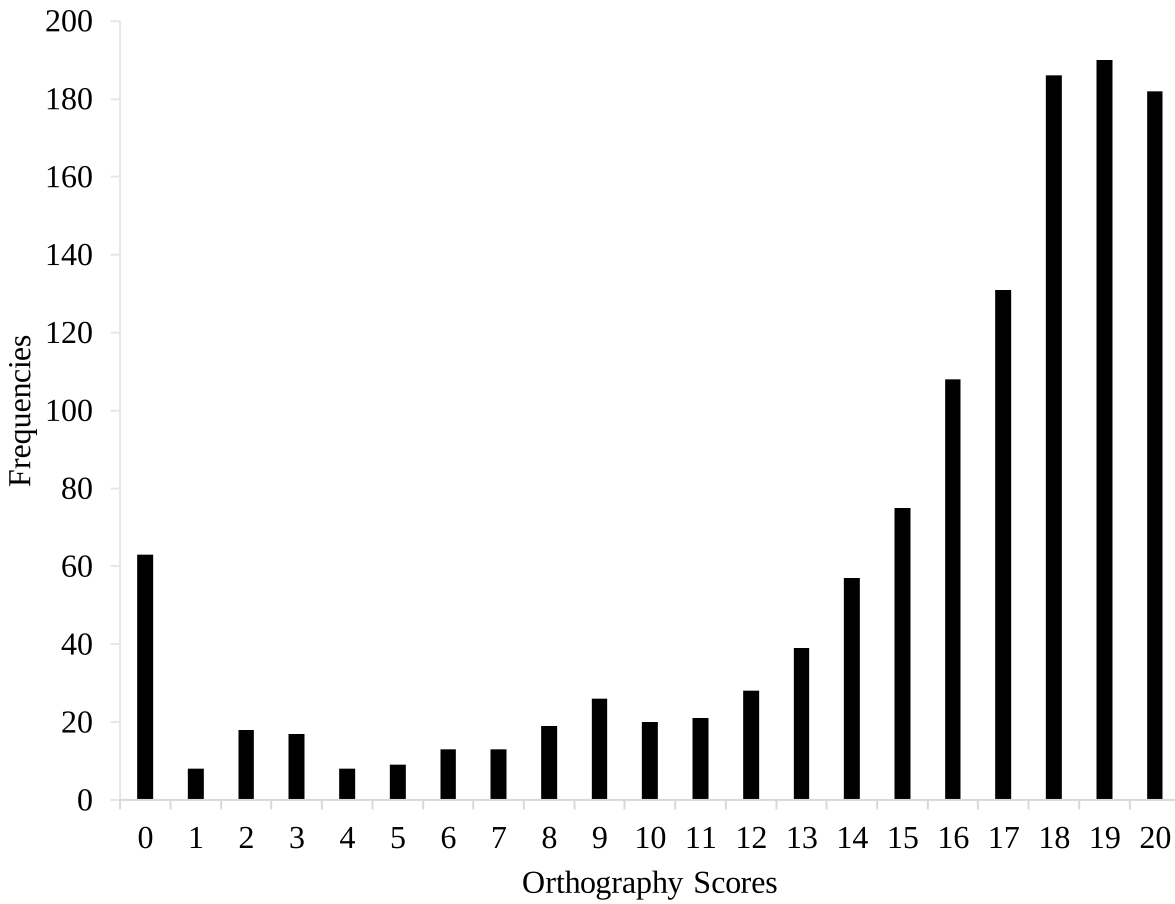


Figure 1. Orthography Score Distribution

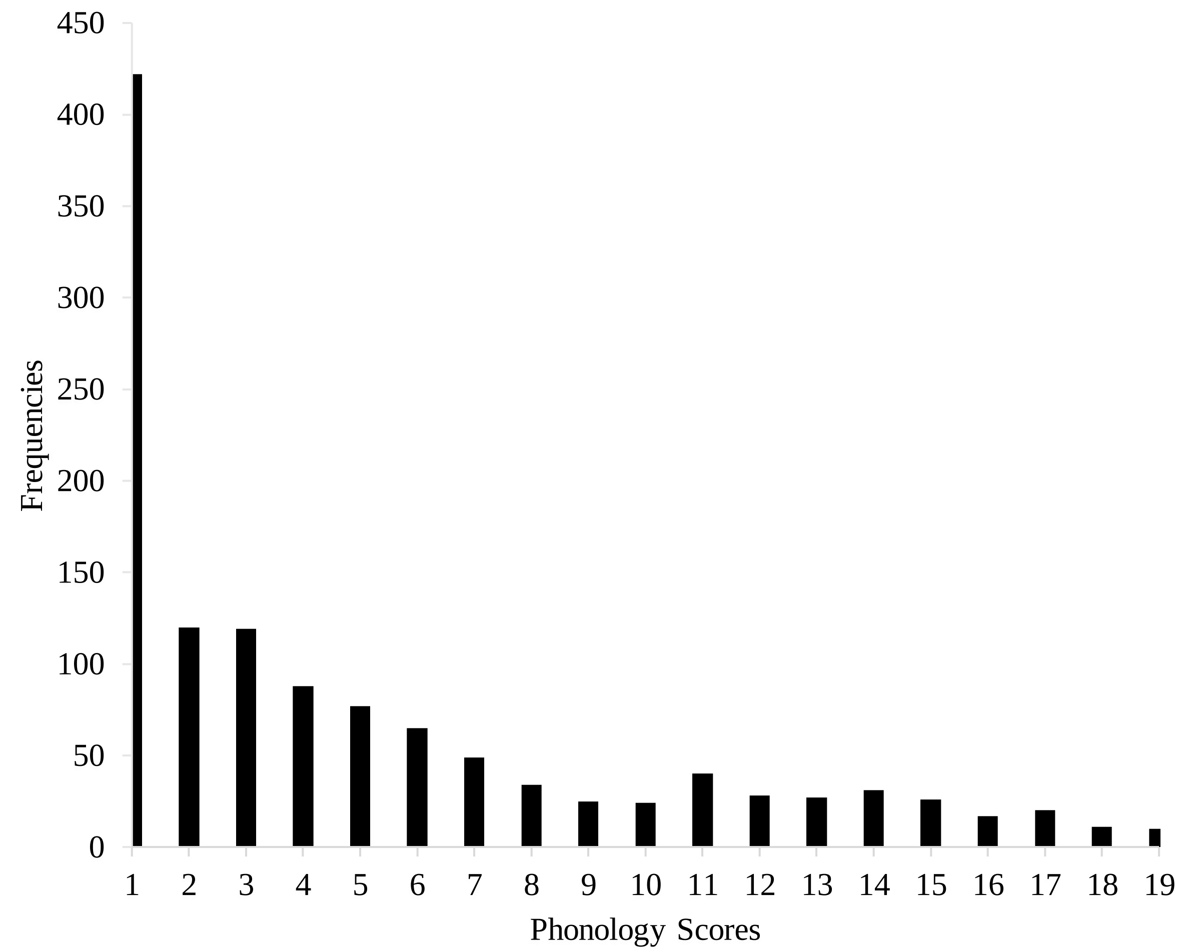


Figure 2. Phonology Score Distributions.

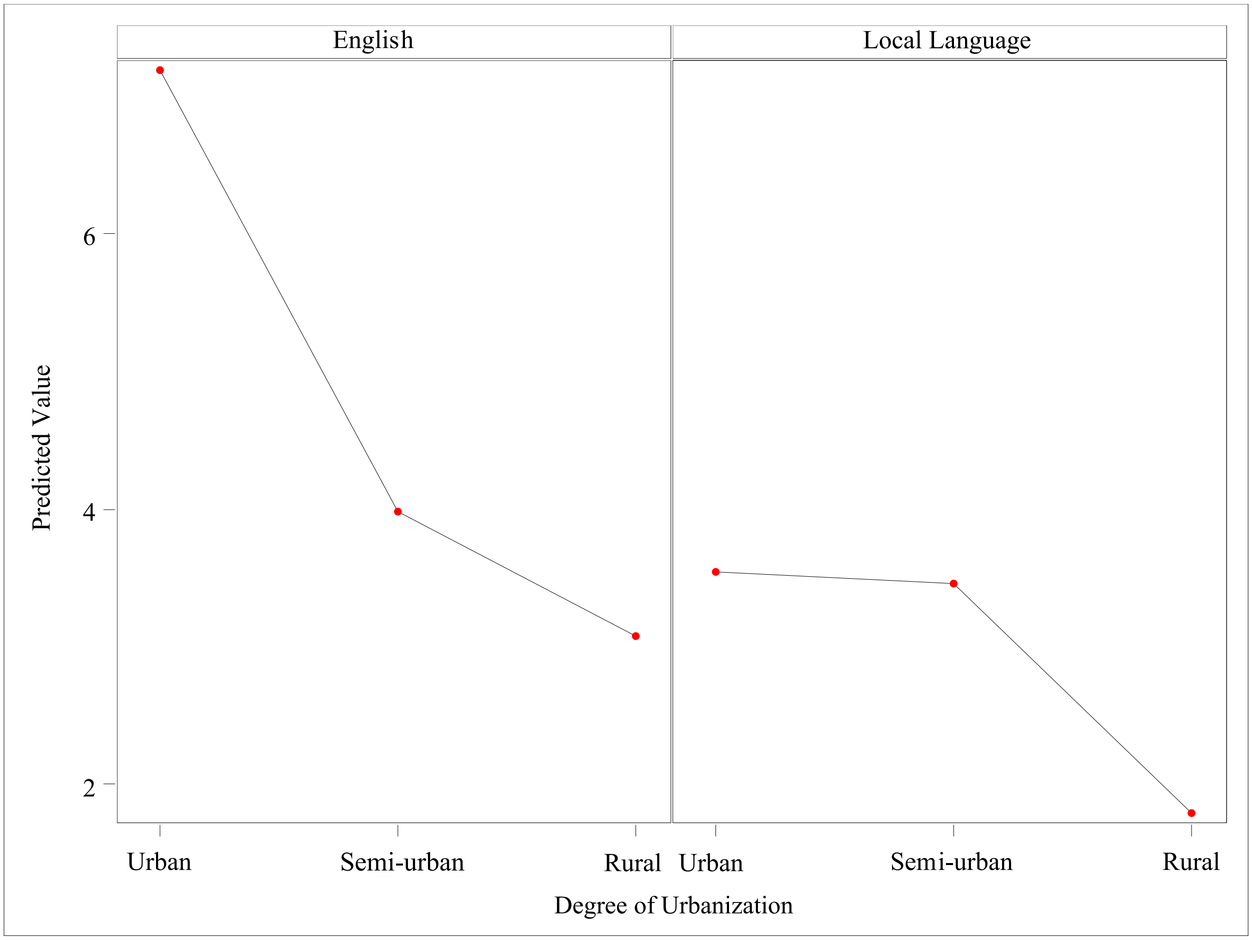


Figure 3. *Effect of degree of urbanization by language of test administration on Phonology Predicted Score Value.*

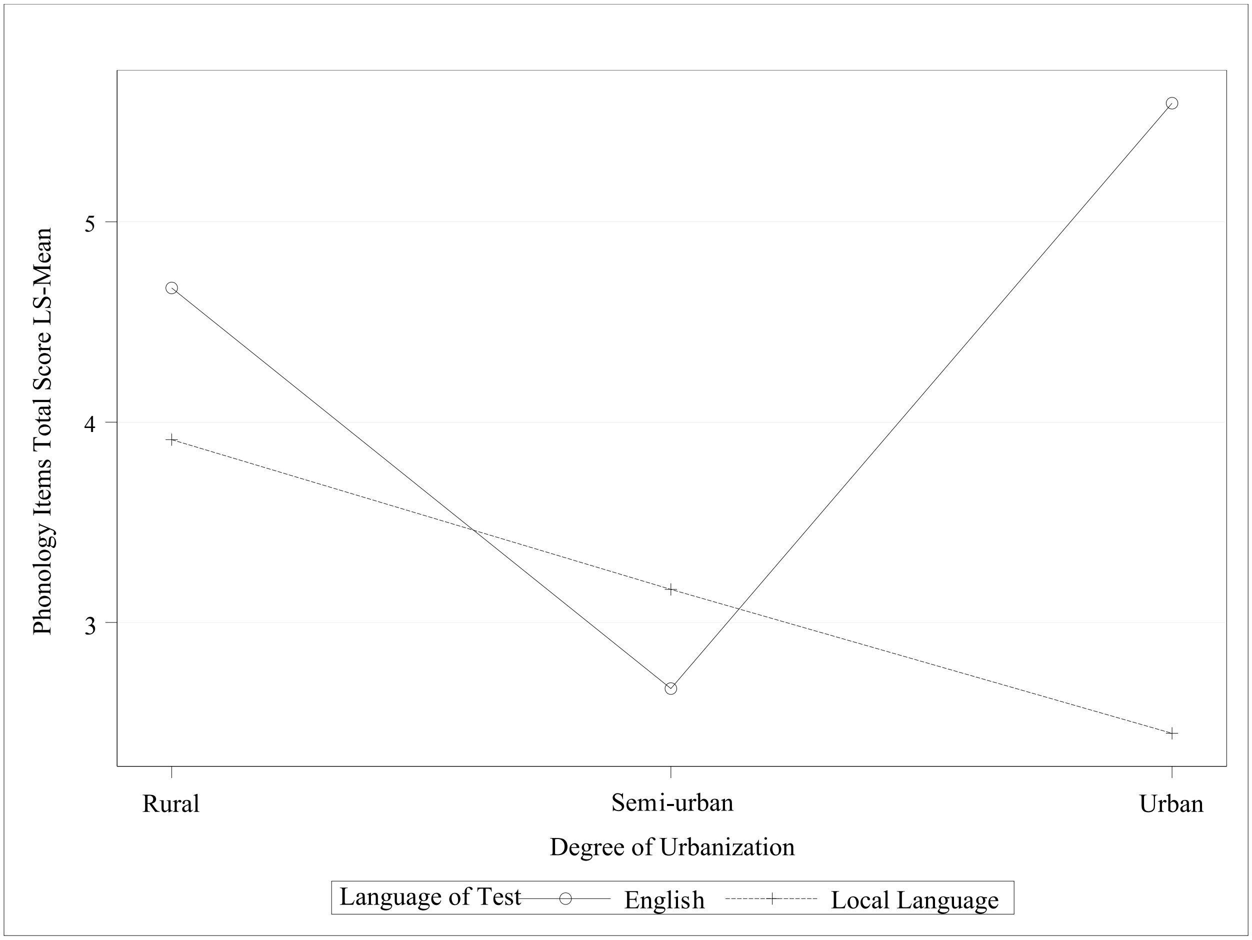


Figure 4. Phonology Interaction Plot for Language of Test Administration and Degree of Urbanization.

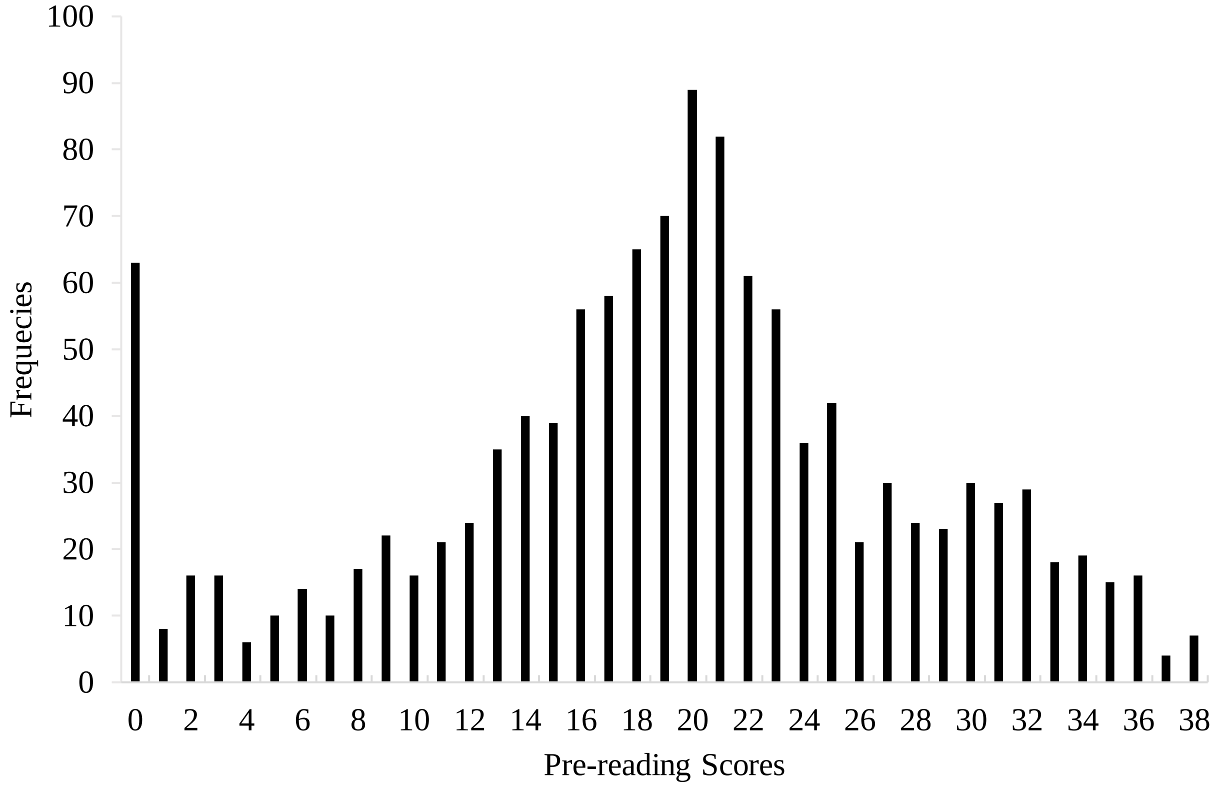


Figure 5. GAT Pre-Reading Score Distribution.

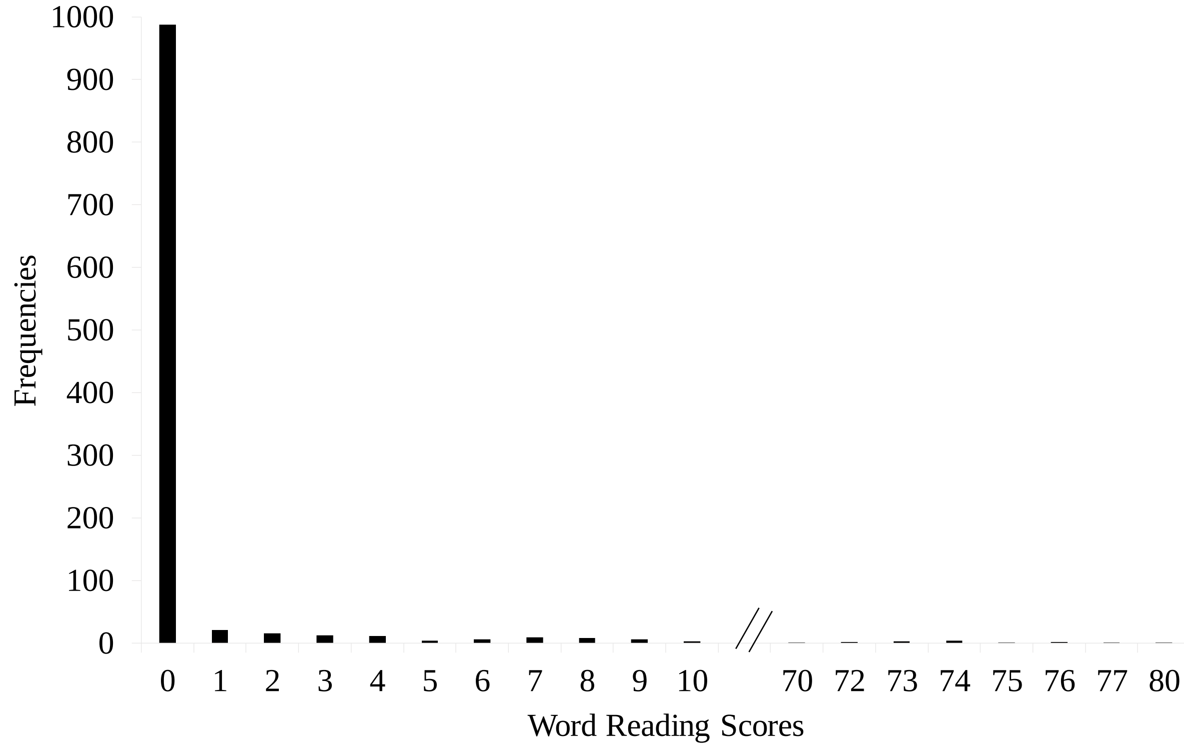


Figure 6. Word Reading Score Distribution.

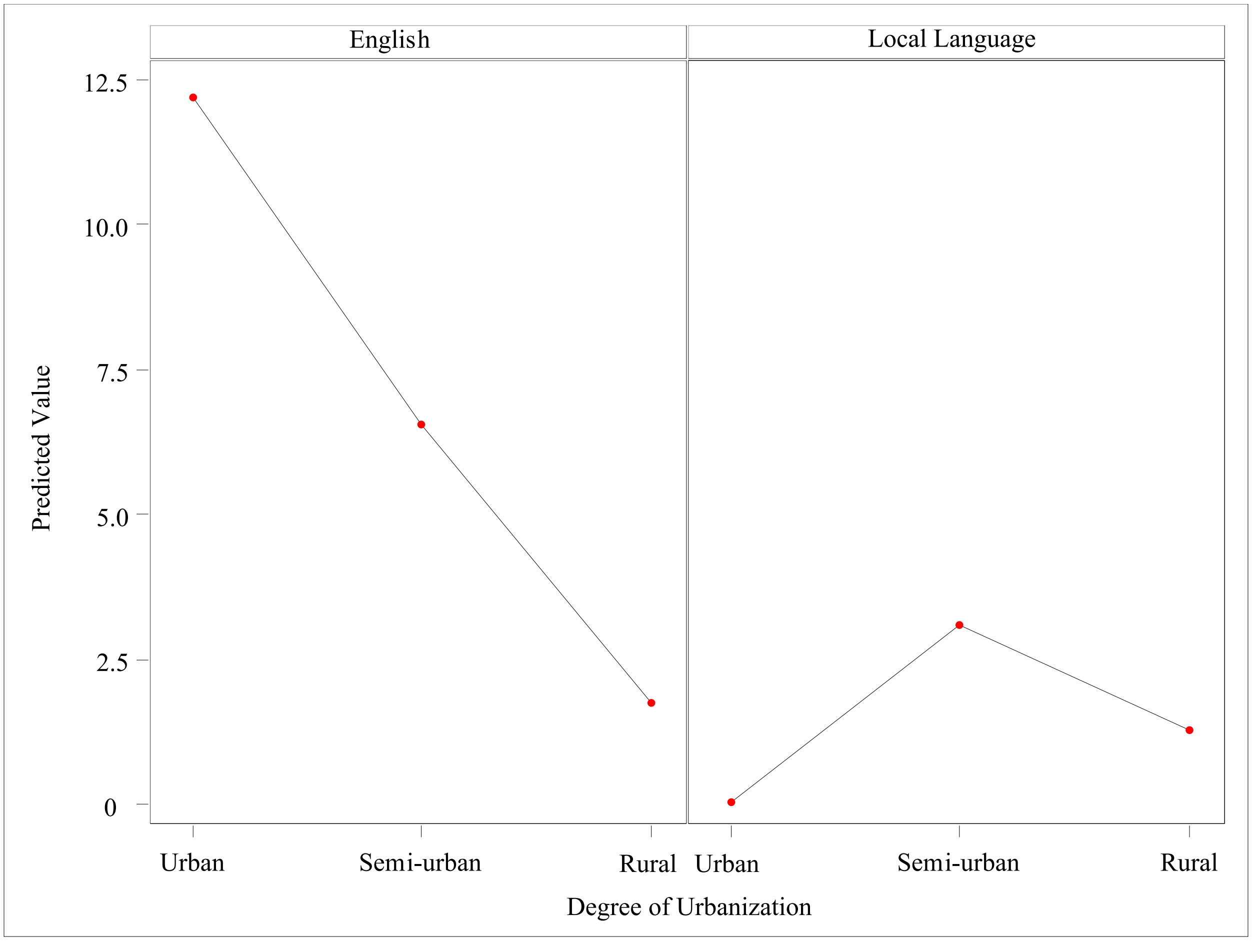


Figure 7. Effect of degree of urbanization by language of test administration on Word Reading Predicted Values.

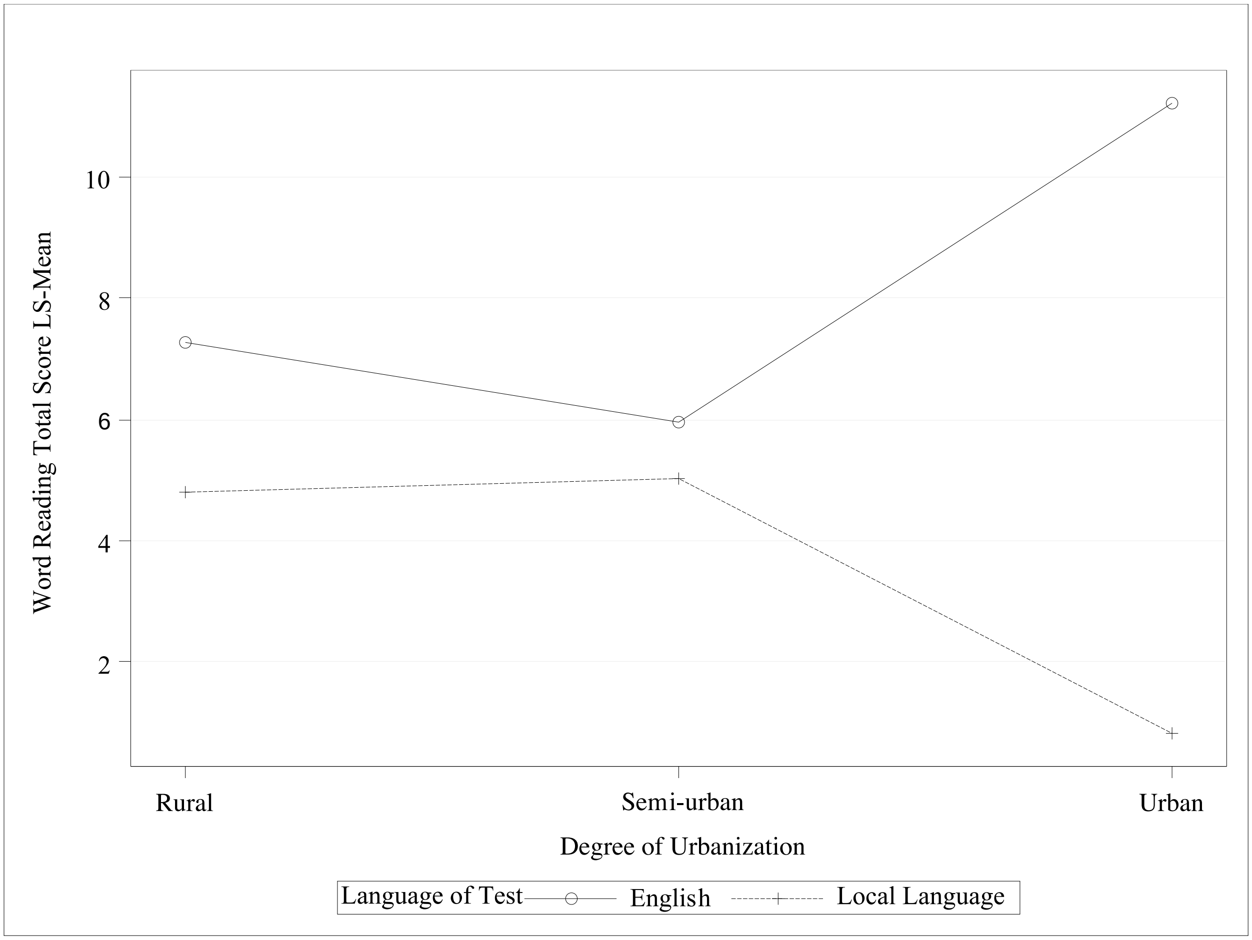


Figure 8. Word Reading Plot for Language of Test Administration and Degree of Urbanization.

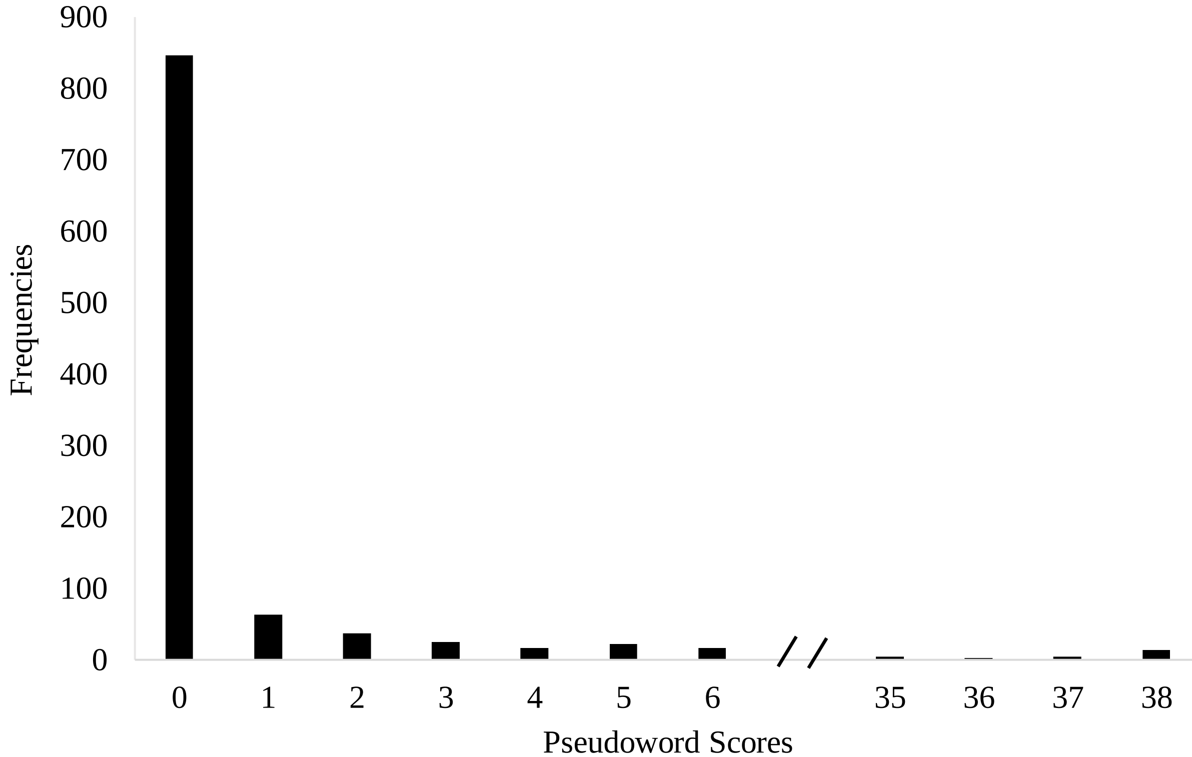


Figure 9. Pseudoword Score Distribution

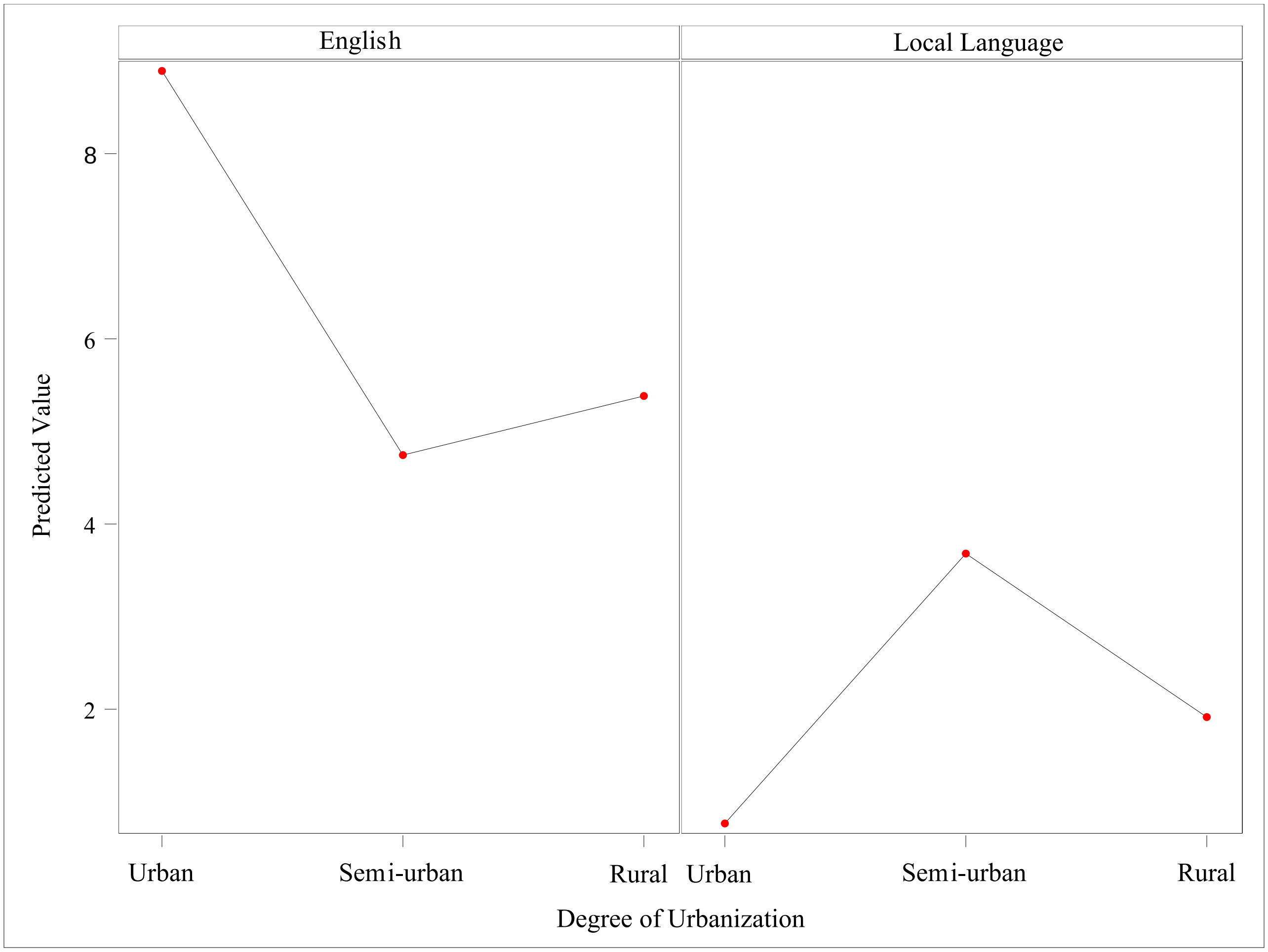


Figure 10. Effect of degree of urbanization by language of test administration on Pseudoword Predicted Values.

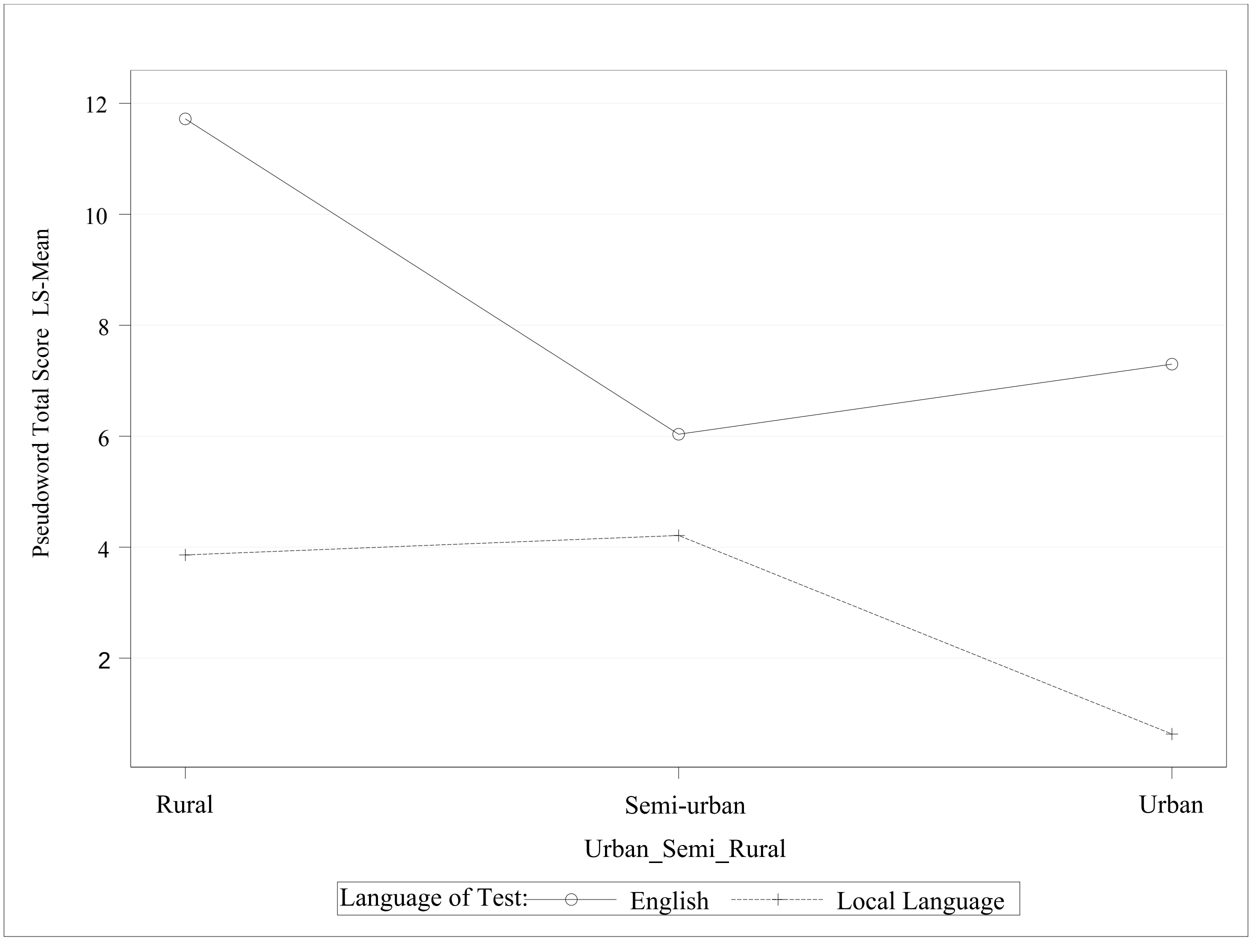


Figure 11. Pseudoword Interaction Plot for Language of Test Administration and Degree of Urbanization.

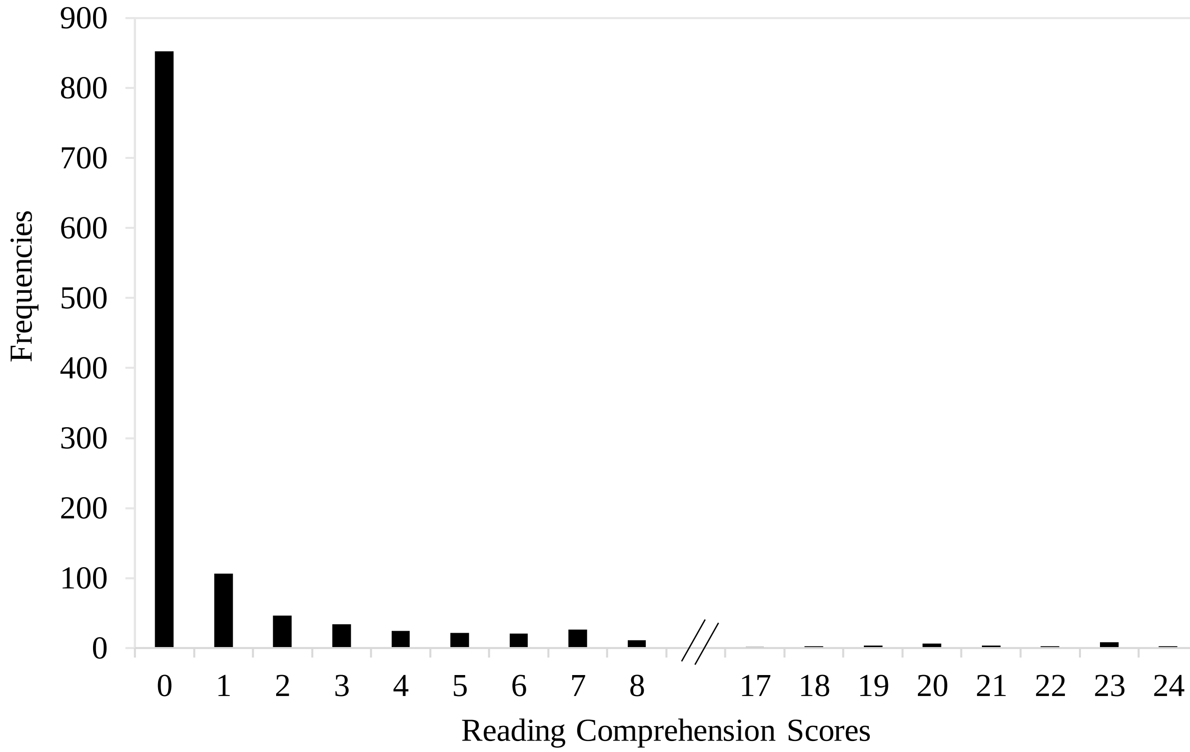


Figure 12. Reading Comprehension Score Distribution

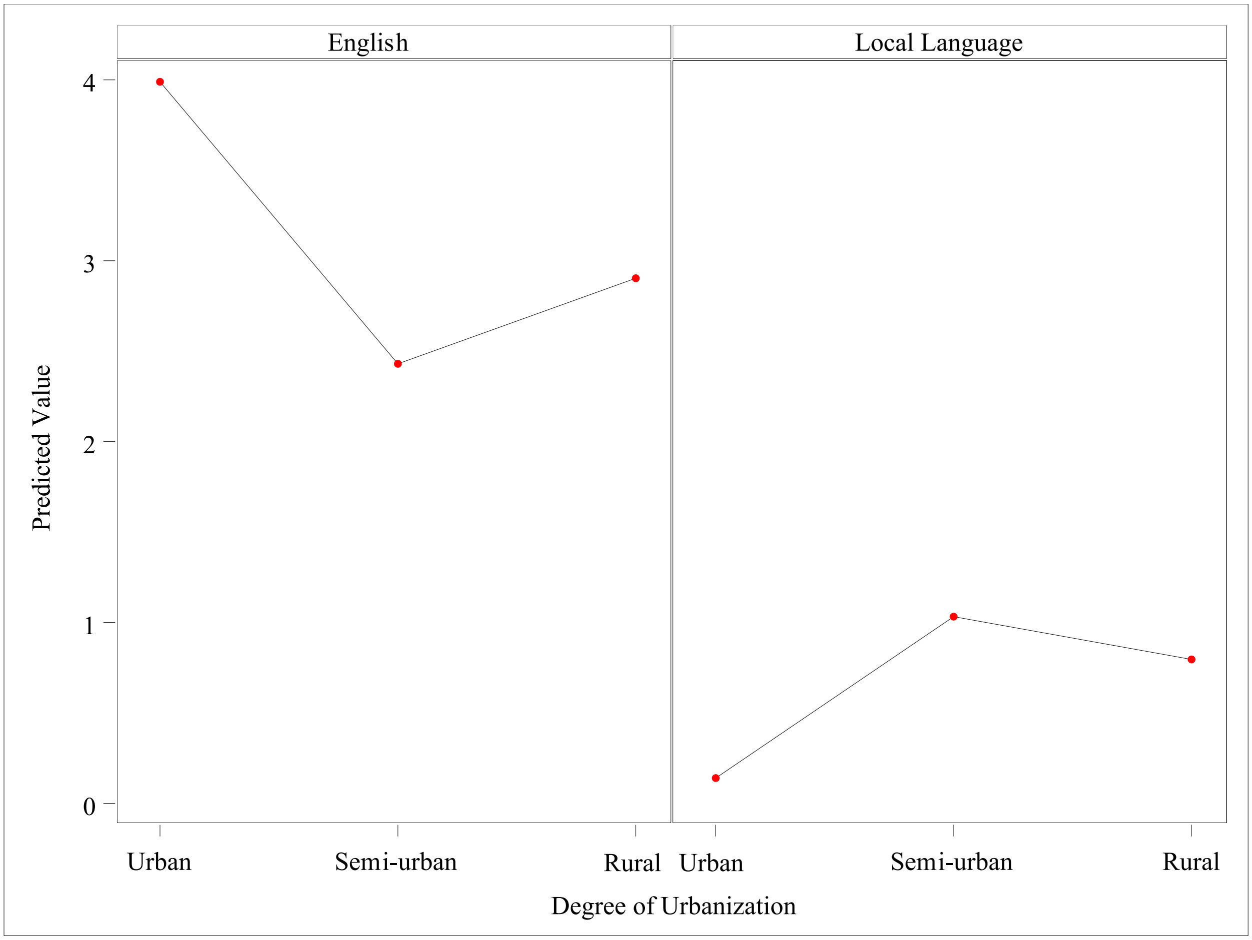


Figure 13. Effect of degree of urbanization by language of test administration Reading Comprehension Predicted Scores.

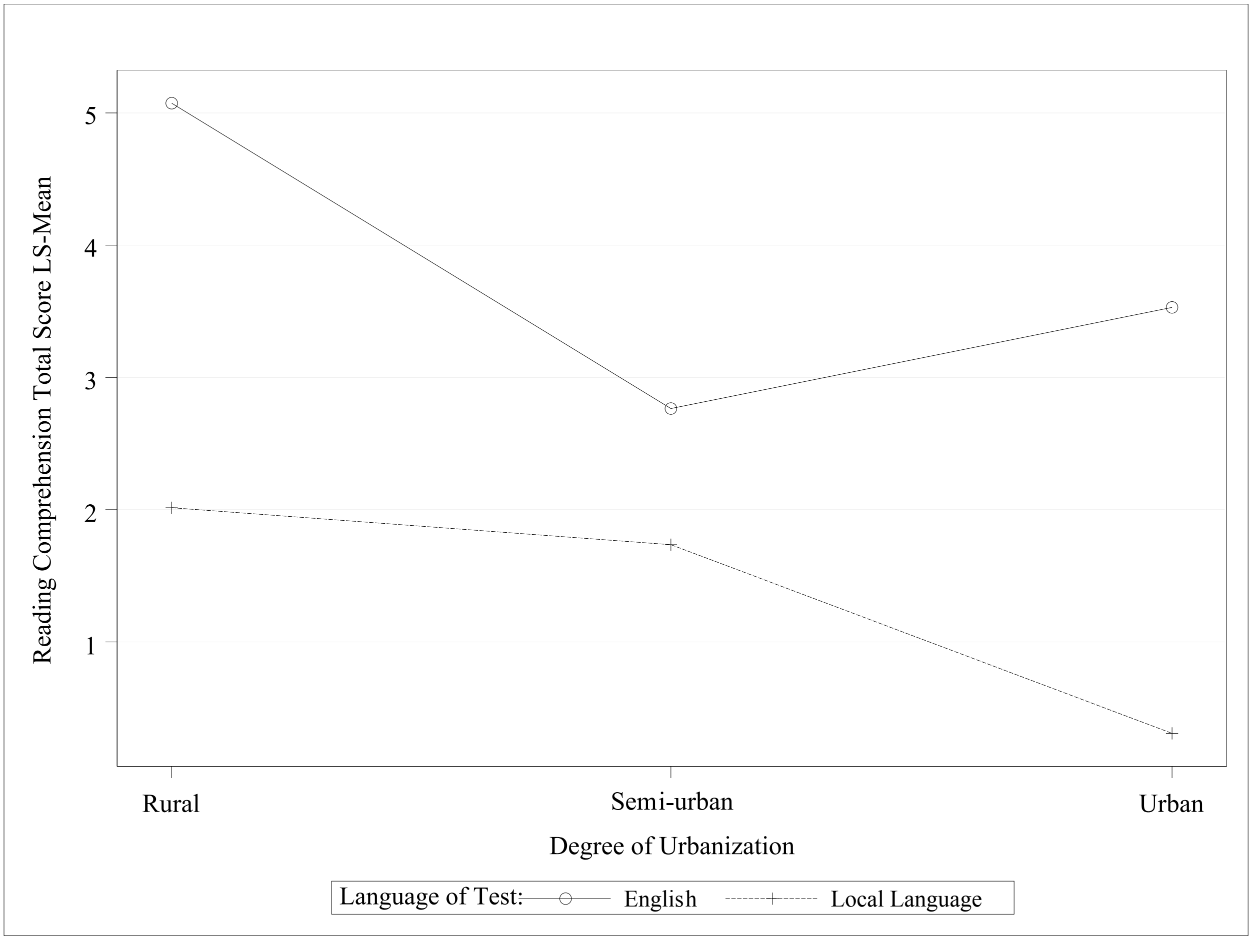
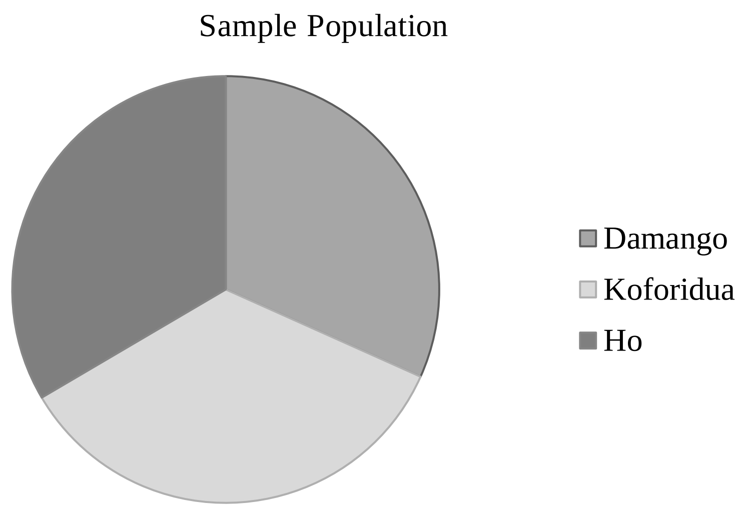
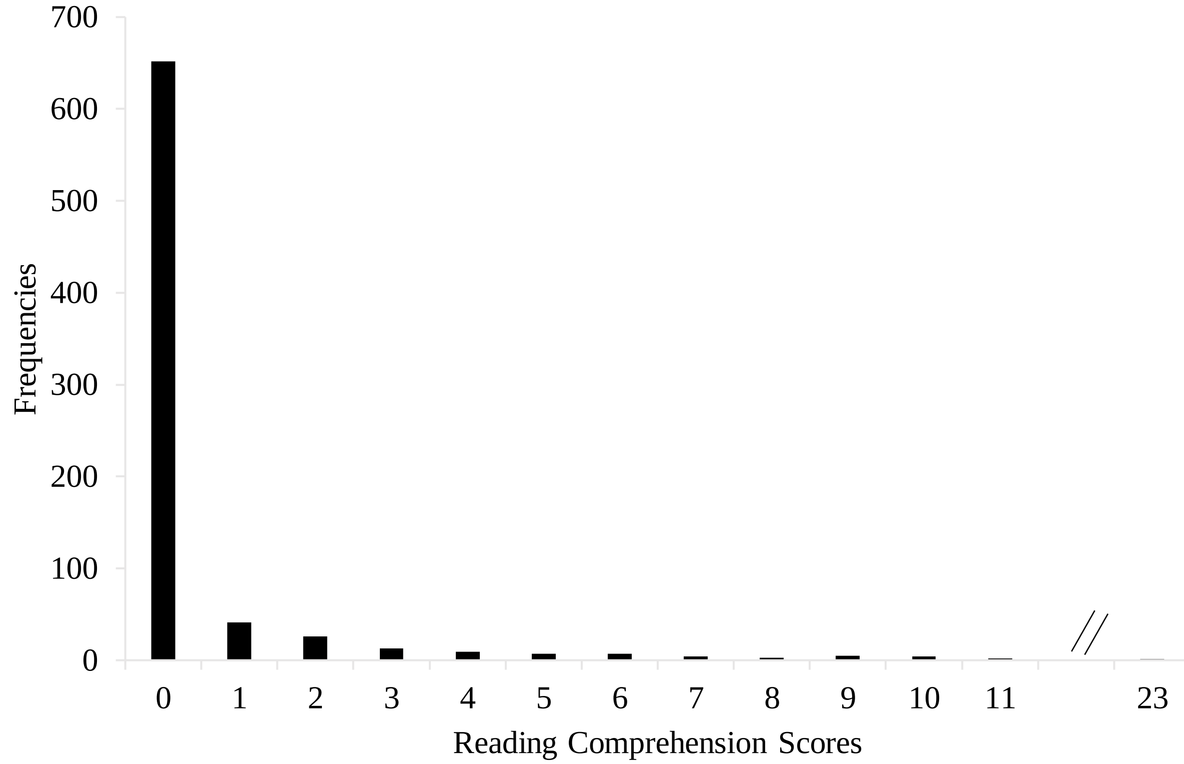


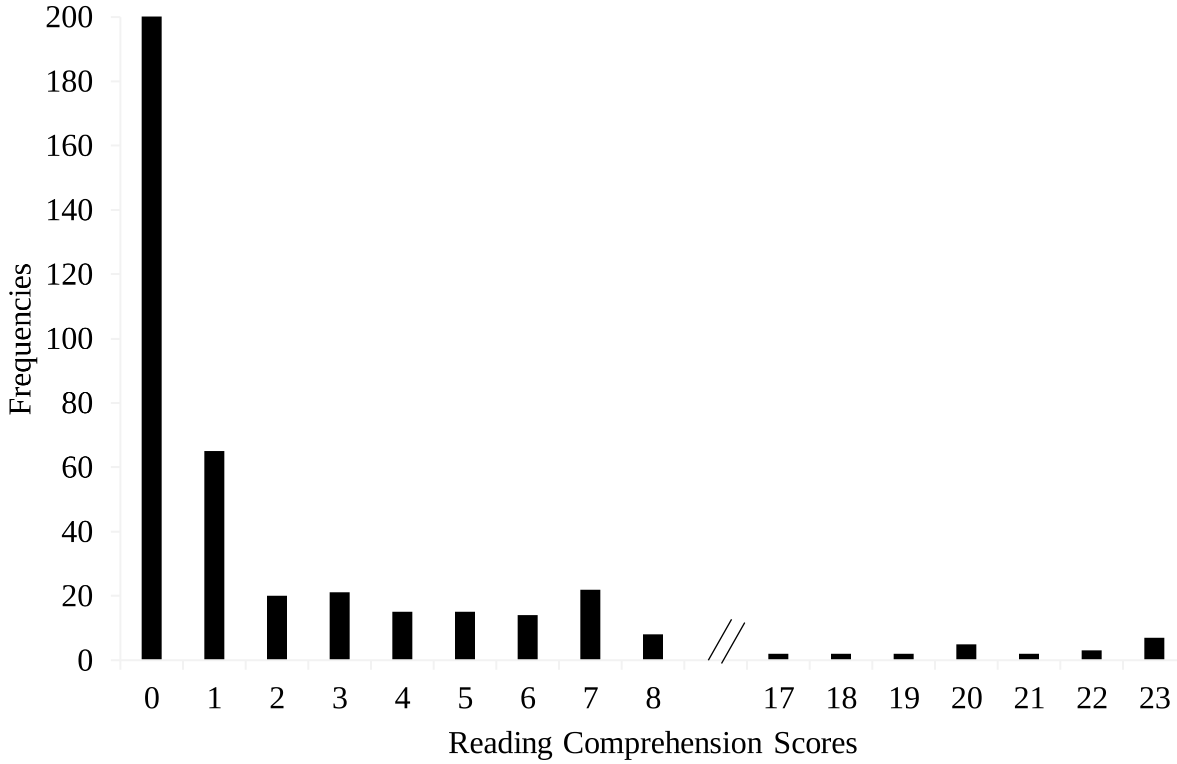
Figure 14. Reading Comprehension Interaction Plot for Language of Test Administration and Degree of Urbanization.



Supplemental Figure 1. Sample Distribution by Region.



Supplemental Figure 2. Distribution of Reading Comprehension (Test Administered in Local Language).



Supplemental Figure 3. Distribution of Reading Comprehension (Test Administered in English).