What can Botswana learn from the TIMSS assessments of 2003?

Poloko Chepete, Molepolole College of Education, Botswana, pclecha@gmail.com

Abstract

This study offers a holistic examination of 8th grade student performance in mathematics in schools in Botswana based on the country's sample data from the IEA's Trends in International Mathematics and Science Study (TIMSS) of 2003. The analyzed sample is comprised of 4,805 students in 136 schools. Missing data was treated at the student level using multiple imputation and at the school level using the expectation maximization (EM) algorithm. Part of the theoretical framework for variable selection was based on six factors from the Rand model (Oakes, 1986). Two of the factors (*student background and attitudes*) were at the student level, three factors (teacher attributes, school quality, and instructional quality) were at the school level, and the sixth one, *Mathematics Achievement*, served as the outcome variable. A two-level hierarchical linear modeling technique was employed to model student achievement in mathematics as a function of the student and the school level factors. Findings suggested that about one fifth of the variability in mathematics performance is between schools and that the learning of mathematics in Botswana was driven largely by student level factors with the attitudinal factor accounting for a greater part of the variation than the student background factor. The findings also suggested that the most effective schools in mathematics had safe school environments and an academically-inspired student body. Teacher-attribute and instructional quality factors served a moderating effect on the relationship between student-level factors and student performance.

Key Words: Missing data, Multiple imputation, attitudes, school safety, hierarchical linear modeling

Introduction

Botswana views TIMSS surveys as nationally representative of the schooling system in the country as well as one of the dependable means through which the country can acquire reliable evaluations of the efficacy of its educational practices from an international perspective. The country hopes to obtain empirical evidence from the analyses of the TIMSS databases that can be used to guide future debates on educational reforms and expenditures. To that effect, Botswana encourages researchers and scholars to engage in secondary analysis of its data and to vigorously explore the data through various analytic techniques. That said, a review of the literature on students' mathematics-achievement in Botswana did not uncover any studies that examined students' achievement holistically as a function of the various components of the education system. Most studies did not deviate much in their statistical methodologies from the analytical techniques used in the TIMSS international reports (Ministry of Education, 2005; Cele, n.d.). These studies mainly reported on the significance of a list of predictor variables in relation to the outcome variable or investigated the performance of single variables under different schooling conditions and not much about how the various components of the education system relate to one another to produce the overall effect.

The current study continued the exploration of Botswana's TIMSS data using other statistical techniques and attempted to answer a multifaceted question: *What can Botswana learn from the TIMSS tests about its education system?* In this study, I explored possible relationships among schools, teachers, and students' contextual variables in relation to mathematics-achievement scores of 8th graders in Botswana, using the TIMSS 2003 online databases (http://timss.org). The main focus of this paper was to model predictor variables that can be manipulated through educational policies such as school environmental conditions and teacher-preparation policies. The paper examined the variance in students' performance according to the hierarchical nature of school data. Through modeling the data I hoped to obtain a global picture of the learning of mathematics in schools in Botswana. My analysis took into account the fact that students' learning achievement was affected by a complex system of factors and that the impact of each of these factors could not be fully understood in isolation but in relation to others (Hiebert, Stigler, Jacobs, Givvin, Garnier, and Smith, 2005; Goldenberg, 2000). The research questions guiding my study were:

(i) What are the student and school level characteristics that combine best to explain students' mathematics achievement in Botswana?

(ii) Which of these variables are amenable to policy interventions?

Modeling of educational data has been in existence for decades; the literature on models of educational factors and their performance as a system is extensive (Glasman, & Biniaminov, 1981; Kaplan & Elliott, 1997; Kaplan & Kreisman, 2000; Koller, Baumert, Clausen, & Hosenfeld, 1999; Oakes, 1986;). That said, models of interest to my study are those that expressed students performance as a function of some resources as well as taking the learning processes into account (e.g. the Rand Model by Shavelson, McDonnell, & Oakes, 1989; and the Education Productivity model by Walberg, 1981 – though mainly suitable for longitudinal data). Specifically, the model that informed my variable selection was the Rand Model (Oakes, 1986; Shavelson, McDonnell, & Oakes, 1989). I chose the Rand model for my conceptual framework because of its scope of coverage of educational indicators. The Rand model presents a holistic conception of student learning in a classroom setting and it appears frequently in the literature that analyzes large-scale data (e.g. Kaplan and Kreisman, 2000; Koller, Baumert, Clausen, & Hosenfeld, 1999 analyzing TIMSS data). It is taken as one of the influential models in shaping public opinion and policy on how to foster school improvement (Kaplan & Kreisman, 2000).

The Rand model is comprised of ten factors grouped under three components of an educational system: *inputs, processes,* and *outputs* (Figure 1).

[Take in Figure 1 about here-Rand Model]

Briefly stated, the model's inputs are the human and financial resources available to education, its processes are what is taught and how it is taught, and its outputs are the consequences of schooling such as academic achievement, participation (what percentage graduate versus drop out), and attitudes (Shavelson, McDonnell, & Oakes, 1989). Though it might not be necessary for every indicator that has ever been suggested for collection to actually appear in the model (Kaplan & Elliott 1997), Oakes (1986) advised that a single indicator of each component of the educational system was inadequate. In his view, "without a series of indicators that assesses all important facets of the schooling processes (the 3 components of the model), we can neither understand the system's overall health nor determine the conditions under which a particular goal is met" (Oakes, 1986, p. 8). What can be deduced from Oakes' remarks is that studies that model only one component of the educational system might not be doing an adequate job of conveying the necessary information about the school effects.

Taking Oakes's remarks into consideration, I utilized the Rand model structure to reach the different components of the education system holistically. In particular, rather than employ all the ten factors of the Rand model, I used only those factors which measures (manifest variables) were shown to correlate significantly ($p < .05$) with student achievement based on the findings from the country's TIMSS report by the Ministry of Education of Botswana (Ministry of Education, 2005) and supported by the literature elsewhere. Specifically, I used six factors out of the ten. The six factors that guided this study were representative of the three components (Inputs–Processes-Outputs) of the Rand model. My six-factor-framework was comprised of: input domain (*student background, teacher attributes, school quality,* and *attitudes factors*); process domain (*instructional quality* – in which I incorporated measures of the teaching-quality factor); and output domain (the factor of *mathematics achievement scores)*. One of the adjustments I made regarded the positioning of the *attitude* factor. Though the Rand model regarded attitude as an outcome variable, I used it in this study as one of the predictor variables mainly because the statistical methodology I used for my analysis, Hierarchical Linear Modeling (HLM), models only one dependent variable as a function of other variables. In addition, the literature about the relationship between attitude and achievement was inconclusive about the direction of the effect. Some researchers believed that attitudes influenced achievement (Broeck, Opdenakker, & Damme, 2005), while others believed the opposite to be true (Shavelson et al. 1989).

Among the factors I omitted from my model-framework were the *fiscal and other resources*, and *curriculum quality*. The measures (e.g. school resources, buildings, and textbooks) pertaining to these factors were found insignificant as predictors of student achievement in the TIMSS report by the Ministry of Education (2005). One of the reasons why these measures might have turned out to be poor predictors of achievement can be due to the homogeneous nature of the education system in the country. Botswana follows a highly centralized education system. School resources are distributed centrally, there is one national curriculum, textbooks are centrally recommended nationwide, school buildings are the same, and most (over 90%)[1] teachers have the same educational background[2]. It would be of little value to expect variation in student performance to be explained by such homogeneous factors.

The six-factor framework used in this study and the indicator variables for the respective factors are depicted in Table 1. Needless to say, the TIMSS databases do not present contextual variables grouped in factors as suggested in Table 1. I extracted the indicator variables based on the literature, from exploring with data, and from the findings based on the report by the Ministry of Education about the TIMSS assessment (Ministry of Education, 2005). Some of the manifest variables were single-item indicators while others were composite (made out of a combination of multiple item-responses). A summary of the six factors and their measures was as outlined.

1. *Student-Background factor*: The measures for this factor were (i) Mother’s educational level – Parents can affect their children’s achievement in several ways such as through hereditary issues (genes) and making opportunities for them to learn (hiring tutors). In the case of Botswana, mother’s educational attainment had a stronger relationship with student achievement than the father’s educational level; father’s education level variable had more missing values than mother’s education. (ii) Books in the home – Studies have shown the variable ‘Number of books in the home’ to be a strong predictor of student achievement in almost all countries (Yang, 2003) including Botswana (Ministry of Education, 2005). (iii) Language of the test – Research has shown that when the official language of instruction differs from the mother’s tongue (or language spoken at home) achievement was negatively affected (Abedi & Lord, 2001). In Botswana, language of instruction is English, which is different from mother’s tongue. (iv)Home possessions: For this index variable I conducted factor analysis for the Botswana’ TIMSS data and the five home possession indicators below loaded relatively high (.571 to .689) under one factor. The items were possession of electricity in the home, television, flushing toilet, refrigerator, and telephone.
2. *Attitudinal-factor*: According to Shavelson, McDonnell, and Oakes (1989), an education indicator system that does not include some attitude indicators would be widely perceived as incomplete. For the current study, I derived measures of attitudes under three composite variables using the students’ responses to the TIMSS questionnaire. The composite measures for this factor were (i) self-confidence, (ii) value math, and (iii) student educational aspiration.

[1] Unit of analysis in TIMSS is the student – so, over 90 percent of the students in the TIMSS sample were taught by teachers who hold a three-year diploma in secondary education (as expected in the system).
[2] Teachers for the TIMSS’ population 2 (Form 1/ grade 8) follow a three-year diploma qualification in secondary education (same math curriculum and number of periods across colleges).

3. *Mathematics achievement factor*: A set of five plausible values given in the databases was used in the analysis (see methodology section).

4. *Teacher- attributes factor*: Measures for this factor were not easy to find; I included some of the measures in the analysis even when they were not significant in the case of Botswana based on the country's TIMSS report, mainly to check on their behavior in the presence of other variables. As noted earlier about the homogeneity of teacher education in Botswana, most of the proxy variables for teacher attributes were not strong correlates of students' achievement. However, the following measures were used (i) Teaching experience - Though Botswana was noted to have the youngest and most inexperienced teachers of mathematics internationally (an average of six years of teaching experience against an international average of 16 years) (Mullis, Martin, Gonzalez, & Chrostowski, 2004), teaching experience was one of the few, if not the only teacher attribute that correlated significantly with students' achievement (Ministry of Education, 2005). The other measures of teacher attributes that were included in the analysis were (ii) teacher certifications and (iii) teachers' beliefs. The measures (ii) to (iii) were not significant predictors of achievement in Botswana but they were internationally. I ran a correlation analysis for the teacher attributes on student achievement using the international databases and noted a Pearson's r = .378 for teaching experience on math achievement and r = .329 for teacher certification on math achievement. The index variable *teacher's beliefs* was very challenging to compose in the case of Botswana. Teachers depicted both traditional and non-traditional beliefs about the learning of mathematics. Seventy-six percent of the teachers did not expect their students to learn mathematics through memorizing but yet believed that mathematics should be learned as a set of algorithms[3] or rules.

5. *School Quality factor*: This factor was estimated using three measures (i) school locality - students in schools in urban areas were found to perform significantly higher than those in rural and semi-urban schools (Mandeville & Liu, 1997; Ministry of Education, 2005) (ii) safe environment (as perceived by students) – It was found that students in schools where there was vandalism, cheating, profanity, and intimidation or verbal abuse of staff (by students) performed significantly lower than the schools where these behaviors were not a problem (Brookover et al. 1978; Ministry of Education, 2005). The other measure was (iii) safe environment (as perceived by the teacher).

6. *Instructional Quality factor*: Measures for this factor included (i) passive learning, and (ii) interactive learning. TIMSS international mathematics report (Mullis, Martin, Gonzalez, & Chrostowski, 2004) revealed that eighty percent of the students in Botswana spent at least half of their lessons practicing numerical operations (+, -, x, ÷); studies show that when students are exposed to rigorous mathematics content, learning increases (Stevenson, Lee, and Stigler, 1986). One wonders if practicing numerical operations can really expose students to rigorous mathematics content that will result in increased learning.

Methodology

Due to the nested nature of classroom data, this study employed a multilevel analytical technique in acknowledgement that students in the same classes are more likely to have a lot in common (teachers, textbooks, share same values, and class culture) and have correlated classroom-responses that lead to correlated errors which violate the assumption of independent errors if standard regression techniques were used. [For a detailed discussion about the effects of analyzing nested data as though single-leveled, see Burstein, 1980; Hox, 1998; Raudenbush & Bryk, 1988-1989].

In brief, while standard regression estimation methods analyze nested observations as though they were obtained from independent or randomly selected cases, multilevel regression techniques estimate the

[3] *Algorithm* is not a very common concept in Botswana. It is possible that teachers misinterpreted it to mean constructive learning or any of the terminologies that indicate active/progressive learning; otherwise, it will be interesting to find out how they (74% of the teachers) expect their students to learn procedurally (mainly through algorithms) without memorizing..

degree of dependence of the responses and incorporate this dependence into the estimation of effects (Raudenbush & Bryk, 1988-1989). Additionally, while ordinary least squares (OLS) regression estimation method make assumptions of homogeneity of slopes (i.e. the assumption that the effects of the covariates on the dependent variable is constant within classes or groups), multilevel regression techniques enable “the investigator to test for homogeneity of regression and provide a sensible way to proceed, regardless of the outcome of this test” (Raudenbush & Bryk, 1988-1989, p. 444). The multilevel analytic method employed in this study is the Hierarchical Linear Modeling technique (HLM), using the HLM statistical package by Raudenbush, Bryk, and Congdon (2005) [version 6.04, August, 2007]. In Hox’s (1995) view, HLM is a hierarchical system of regression equations that takes into account the interactions of variables between levels in addition to appropriately accounting for the use of two different degrees of freedom (one regarding the number of schools and the other regarding the number of students). Hox (among others) contends that failing to adjust for clustered data produces misleading significant results (spuriously significant effects).

This study used secondary data from the international comparative assessment TIMSS (2003), 8th grade databases available online (www.timss.org). The Botswana TIMSS sample was comprised of 146 schools, 146 teachers, and 5,150 students. This study was reporting on the subject of mathematics. Like most large-scale surveys, Botswana’s TIMSS databases were invested with missing data especially at the school level where about seven percent of the teachers did not respond to the questionnaires. Consequently, hundreds of the students from those schools had missing data at the school level. Actually, every background variable used in this study had missing values [except for the aggregated variables because they assume mean scores). The percentage of missing data among the variables ranged from a high of 14.6 (*student educational aspiration*) to a low of .7 (frequency of speaking English at home) among the student-level data, and from a high of 11 (teaching experience)[4] to .7 (teacher educational attainment) among school level variables. Moreover, missing data in this study could not be assumed to be missing completely at random (MCAR). In my exploratory data analysis, students with missing data performed significantly lower than those with available data in almost all the variables. Deleting missing cases in this research would have biased parameter-estimates. Besides, the APA Task Force described the methods of deletion of missing values as the worst methods available for practical applications (Wilkinson & The Task Force on Statistical Inference, 1999). Upon a review of the literature on treatment of missing data, maximum likelihood substitution methods were found to be superior to the traditional missing data handling methods such as mean substitution methods including substitution based on predicted values derived from regression lines (Tabachnick & Fidell, 2001; Allison, 2002; Peng, Harwell, Liou, & Ehman, 2006; Schafer & Graham, 2002). Specifically, methods that use iterative processes such as the expectation maximization (EM) algorithm, full information maximum likelihood (FIML), and multiple-imputation (MI) were recommended in the literature, with MI said to perform well under any missing data pattern (whether missing at random or not) (Tabachnick & Fidell, 2001). Needless to say, missing data issues are more complex in multilevel data structure than in uni-level data. Gibson and Olejnik (2003) had issues with the use of multiple-imputation at a group-level in nested data. According to them, MI distorts estimates at level-2. The two found the EM algorithm to be more reliable for treating missing data at level-2, but were silent about the use of MI at level-1.

That said, after eliminating cases with missing data in all response categories[5] I replaced missing data at the student level (level-1) using Multiple-Imputation through SAS software (version 9.1.3, 2007). Ten datasets of imputed values were generated at level-1 [the SAS program[6] used for the multiple-

[4] The response “I don’t know” was counted as missing data in this study and was common for the variables *student educational aspiration* and *mother educational attainment*.

[5] TIMSS technical report requires that at least two-thirds of the responses should be valid for use in the analysis (Ramirez & Arora, 2004).

[6] SAS program: `PROC MI DATA=missing.For_mi OUT=missing.replaced NIMPUTE=10 SEED=13579 MINIMUM = 1 1 1 1 1 1 1 MAXIMUM = 3 5 3 5 3 3 8 ROUND=1;Title`

imputation process is given below in the footnote section]. I replaced missing values at level-2 using the EM algorithm utilizing the SAS statistical software for the program. When imputing level-2 missing data, level-1 variables were not included in the estimation process because that would have resulted in unique mean-values estimated for each student instead of for each school (See Gibson & Olejnik, 2003). Thus, only level-2 data files were subjected to the EM algorithm. In preparing data for HLM analysis, ten multivariate data matrix (MDM) files were created according to the ten multiple imputed level-1 datasets. The final sample consisted of 4 805 students clustered in 136 schools. The class sizes ranged from 8 to 45 and averaged 37 students per class. The sample was such that only one class was sampled from each school. Put another way, the number of teachers in the sample was equal to the number of sampled schools (there was no nesting of teachers in schools). To that end, my analyses used a two-level HLM algorithm with the student variables at the first level and the school variables at the second level.

In regard to the variables, though most measures were selected based on the findings from the country's TIMSS report (Ministry of Education, 2005), in addition to conducting exploratory factor analyses (with Principal Components analysis and Varimax rotation) to screen items for the composite variable, I conducted reliability measures for the scales of the composite variables. I used Cronbach's alpha (a) to measure the internal consistency of the scales. The a –values are given in Table 1 for the composite variables together with other descriptive statistics.

The scores on the variables were such that they reflected desired conditions. That is, the direction of influence was in the direction of performance. There were seven variables (grouped under two factors) tested at the student-level and 15 [eight school-level characteristics (grouped under three factors) and seven aggregated level-1 variables] at the school-level. All variables used in this study were grand-mean centered. Data at the student-level were weighted using the sample weight *Total Student Weight* (TOTWGT[7]), and the weighting variable at the school-level was *Mathematics Teacher Weight* (MATWGT). In relation to the dependent variables, all the five plausible values were used in a single run. HLM scientific software (version 6.04) allows the use of all plausible values and up to ten imputed data files in a single run of an analysis [See HLM 6 manual, Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2004]. It also accommodates weighting variables. As stated by Raudenbush, Bryk, Cheong, Congdon, and du Toit, when multiple imputed values are used in HLM, the program runs each of the imputed datasets internally, produces their averages, and calculates appropriate measurement errors.

The reported models in this study were estimated using the restricted (or residual) maximum likelihood. A one-way random effects ANOVA (empty model) was conducted to provide a measure of the variances within and between schools. This was mainly to answer three basic questions of whether the data analyzed warranted a multilevel procedure; whether schools differed in their mean achievement in mathematics; and how much of this variance in mathematics scores was due to differences among students and how much was due to differences among schools. To facilitate interpretation of the effects of the five factors (*student-background, attitudinal measure*, *teacher-attribute*, *school-quality*, and *instructional-quality*) on the sixth factor, *mathematical achievement*, separate HLM were developed for each of the factors before composite models were developed (Bryk & Thum 1989; Bryk & Raudenbush, 1992). Each of the two student level factors of *student-background* (with four indicator variables), and *attitudinal measures* (with three indicator variables) was entered into the empty model separately to assess its absolute effect. This was followed by adding all the seven student-level characteristics

```
"Imputation outcome";VAR English BsbgbookHomeposs aspiration Selfconf
Valuemath mother;RUN;
```

[7]Total Student Weight (TOTWGT) is used to ensure that the various subgroups that constitute the country's sample are properly and proportionally represented in the computation of population estimates for that country (Foy & Joncas, 2003).

together into the empty model disregarding their grouping factors. This step was carried out to determine whether each variable had a significant relative effect on student achievement in the presence of other variables. The seven student-level characteristics were: *frequency of speaking English at home, books at home, home-possessions, mother educational attainment, student educational aspiration, self-confidence in math,* and *how students value math*. The research question of interest in this section was *what are the student-level characteristics that explain student achievement in mathematics in Botswana?*

To assess the overall effect of each of the school-level factors on school performance, the models were adjusted for student-level differences before developing separate HLMs for the school-level factors (Bryk & Thum 1989). A summary of the results of the HLMs for the five factors is given in Table 1 together with the variable properties

[Take in Table 1 about here –contains factors, variables, and HLMs for factors]

To fully answer the research question, w*hat are the school level characteristics that help explain why some schools have higher means than others,* aggregated student-level measures were also added to the models. Aggregated measures here refer to the school aggregate (mean-values) of the student-level variables. These were added to determine the influence of the student body (or peer-effect) on the individual student performance after controlling for the student-level variable effect. The 15 school-level variables comprised of *safe environment as perceived by the student*, *safe environment as perceived by the teacher, school locality* (the school-quality factor); *teaching experience, teacher qualification*, and *teacher beliefs* (the teacher-attribute factor); *degree of passive learning*, *degree of interactive learning,* (instructional-quality), and the seven aggregated measures *Mean English* (mean usage of English), *Mean Books* (mean books in the home), *mean-home-possessions*, *mean-mother-educational attainment*, *mean-aspiration*, *mean-self-confidence*, and *mean-value-math*.

Findings and Discussion

The results of the one-way random effects ANOVA indicated that about 21% ($\hat{\rho} \approx .207$) of the overall model variance can be attributed to differences among schools that students enroll in while about 79% was due to differences among the students themselves. This .207 (intraclass correlation coefficient – ICC) also gave the estimated correlation between pairs of scores within the 136 schools analyzed in this study that would have violated the standard regression assumption that students' errors are independent (Rowan, Raudenbush, & Kang, 1991). This variance in school mean performance was not trivial as indicated by the results of the null hypothesis H_0: $\tau_{00} = 0$, for the ANOVA test of differences among school means. The chi-square test indicated a significant variation ($\chi^2 = 1246.1$, $df = 135$, $p < .001$). Thus, schools in Botswana were not that homogeneous in their performance even though the country followed a centralized education system.

When the *student-background factor* entered the model, the total variance in mathematics achievement was reduced by *10.2* percent (from a total of 5124.67 to 4601.95) with three of the indicator variables significant at $p < .001$ and mother educational attainment significant at p < .01. The *attitudinal factor* reduced the total variance in mathematics achievement by *19.5* percent (from 5124.67 to 4124.2). All the three indicator variables comprising this factor were highly significant ($p < .001$). Comparing the two factors, the attitudinal factor had the strongest effect (19.5% versus 10.2%). When all the seven student-level characteristics were added together in a fix-coefficient model, the student-level model accounted for 23.9% of the unexplained total variance in math achievement (a drop from a total of 5124.67 to 3899.16). Out of the seven student-level measures, six were statistically different from zero except mother's educational attainment (γ = 1.37, *t* = 1.82, *p* = .07. The most significant student-level predictor was the *student educational aspiration* (γ_{51} = 11.64; *t* = 16.38). The effect of *mother educational attainment* seemed to disappear when considered with other student-level variables. This variable was retained in the model to examine its effect across schools. When all the seven level-1 variables were entered into the empty model with varying slopes, the effect of *mother educational attainment* on student performance remained non-significant (γ = 1.21, *t* = 1.61, *p* = .12) and had no random effect across school either (u4 = 11.84, χ^2 = 151.14, df = 130, p = .099). Actually, only three variables (out of the seven) had effects that varied randomly across schools in this model. Said differently, heterogeneity of regression slopes existed for the variables *number of books* (u_2 = 41.9, χ^2 = 178.81, *df* = 130, *p* <.01); *home possessions* (u_3 = 63.82, χ^2 = 165.78, *df* = 130, *p* < *.05*) and *self confidence* (u_5 = 74.2, χ^2 = 162.19, *df* = 130, *p* < .05). The variable *mother educational attainment* was dropped from subsequent models.

In relation to the effects of the school-level variables, when separate HLM models were run for each of the three factors, the school-quality factor explained about 28.7 percent of the variability among the adjusted school mean achievement with the indicator variables *safe environment as perceived by the students* and *school-locality* showing a significant influence on school mean achievement while *safe environment as perceived by the teacher* portrayed a non-significant effect (γ = 8.29, *t* = 1.57, *p* = .120) (Table 1). The teacher-attribute factor was the second strongest predictor of the adjusted school mean achievement scores. It's model explained about 16 percent of the variance. The effects of the indicator variables *teaching experience* and *teacher qualification* were marginal (Table 1) while the effect of *teacher-beliefs* on student achievement was non-significant. The effect of the third factor, *instructional-quality* was negligible. It accounted for less than one percent (.02%) of the variation in the adjusted school mean achievement and none of its indicator variables were statistically significant. When the three factors (eight variables) were tested for a combined effect together with the significant effects from the student-level variables, the results suggested that school mean achievement was high in schools with safer environment (less abusive language/less theft/less violence) as perceived by the students (γ = 9.2, *t* = 3.24; *p* < .01) and in schools in urban areas (γ = 8.29, *t* = 2.99, *p* < .01). The main-effect of *teaching-experience* was weak as a predictor of school mean achievement (γ = .77, *t* = 1.79, *p* = .08). Three cross-level interactions were observed between *home-possessions* and *teaching experience (*γ = .9, *t* = 2.7, *p* < .01); *student educational aspiration* and *school locality* (γ = 2.73, *t* = 3.5, *p* < .01); and *self-confidence* and *degree of passive learning* (γ = 8.6, *t* = 2.13, *p* < .05).

Finally, the aggregated measures were added to the model after retaining only the significant effects from previous models. Worthy of note from the results of the full model was that aggregated variables seemed to explain away the effect of *school-locality*. In particular, in the presence of the variable *mean-aspiration* (which is the average of student educational aspiration variable), the effect of *school-locality* ceased to exist both as a main effect (predictor of the school mean achievement) and as a moderating variable (cross-level interaction effect with the student-level variable *aspiration*). From exploring the data, the bivariate relations had indicated a moderate-to-high correlation between *school-locality* and *mean-aspiration* (r = .68). The implication here was that these variables conveyed or measured related information. Apparently, students in urban schools had higher educational aspirations than those in schools elsewhere. The general educational aspiration of the school population (peer-effect) explained away the *school-locality* effect as a predictor of school mean achievement.

The findings from the final model reiterated the results from the previous models, mainly that six student-level characteristics had significant effects on students' performance in mathematics and that after controlling for level-1 predictors, school mean achievement scores were high in schools where students felt safe (safe conditions or high school climate) and where students aspired to reach higher educationally. Three cross-level interactions were noted. The contextual effect for averaged home possessions (*mean-*

home-possessions) on the slope of *the number of books* was significant. Thus, the differentiating effect of the *number of books* within a school was dependent on the *mean-home possessions* of the school community; *home-possessions* (student-level) and *teaching experience* (school-level) was another interaction; and *self-confidence in mathematics* (student-level) and the *degree of passive learning* (school-level) was the third interaction. The effects of the *number of books* and *student self-confidence in mathematics* varied randomly across schools even after accounting for the school level variables. The final model accounted for 77.7 percent of the initial variance that was observed in the school mean achievement.

In summary, the findings from this study indicated that schools accounted for about 21 percent of the variance in students' mathematics achievement scores. This was not unexpected given the homogeneous nature (centralized-system) of the country's education system. The schools were bound to resemble each other more than the students they enrolled. This also is inline with findings elsewhere. A study by Koretz, McCaffrey, & Sullivan (2001) indicated that Japan and Korea for instance, each of which follows a national curriculum had percentage of score-variance due to differences between schools (ICC) as low as 8 and 6 respectively while countries that do not have common school curriculum such as Australia and the USA had higher ICC (47% for Australia and 42% for USA).

In answer to the research question: *what are the student- and school-level characteristics that combine best to explain students' mathematics achievement in Botswana?* The results indicate that the education system in Botswana was driven largely by individual-level factors. Specifically, the student attitudinal factor stood out as more crucial to achievement than did the student background factor. This finding added to the literature that claimed positive attitudes influence student achievement (Broeck, Opdenakker, & Damme 2005). Alongside this literature, however, is a dissenting body of research studies that suggests that the high association between attitudes and performance could be a reflection of the unchallenging nature of the curriculum pursued (Ma & Wilkins, 2003; Mullis, Martin, Gonzalez, & Chrostowski, 2004; Schulz, 2005). The findings from the current study cannot dispute or agree with the research that brings the quality of the curriculum into perspective because curriculum aspects were not examined in this study. Nevertheless, since negative (or less positive) attitudes might not improve the quality of the curriculum per se, it is only a positive thing that this study associated positive attitudes with high achievement.

One finding that did not fit existing research literature was the weak effect of *mother educational attainment* as a predictor of student achievement. I found the non-significant result of this variable inconsistent with both the findings from the Ministry of Education report and other studies (Campbell, Hombo, and Mazzeo, 2001). Nevertheless, a more or less similar result of a weak predictive power of parental educational attainment was noted in a study by Koretz, McCaffrey and Sullivan (2001) when modeling TIMSS data of 1995. The researchers regarded the result anomalous and cautioned readers against the general risk of basing conclusions on a single data source. In relation to the findings by the Ministry of Education report in which this variable was significant, I noted that when a separate model of achievement was constructed containing only the indicator variables for the student background factor, *mother educational attainment* had a significant effect (Table 1). However, when modeled with other variables, its effect was non significant. My hunch is that the effect of mother's educational attainment gets explained away by the joint effects of other variables collinear with it. The effect of this variable in model form warrants further investigations.

The findings at the school level indicated that schools with a positive climate (safer and friendlier environment) and an academically inspired student body achieved better. In regard to safe environment, the literature showed that schools with lower levels of violence, less theft rate, and less frequency of abusive language provided better learning environments for students (Gronna & Chin-Chance, 1999). The argument according to Gronna and Chin-Chance was that students in schools with less safe environments had less time to focus on academic work as they worried about other things such as victimization; in addition, some of the teaching time may be diverted to addressing disciplinary issues while in schools with safer environment there would be fewer incidents that interrupted the academic process.

Although the effect of school locality was a good predictor of school achievement, and supported by literature (Mandeville & Liu, 1997; Ministry of Education, 2005) it is on a happier note that it got explained away by one of the aggregated student-level variables, mean-school aspiration because it is not amenable to policy interventions. Put another way, it would be unreasonable to build all schools in urban areas or to provide transport for pupils in rural areas to attend schools in urban schools as a solution to their underachievement. Probably it would be more meaningful to learn about other characteristics of such schools that could inform policy. *Mean-school aspiration for educational attainment is* one such a characteristic. It will be relatively easier to instill a desire among students to learn more, persevere more in mathematics activities and aspire for more mathematics than take all schools to urban areas. Safety conditions in schools in Botswana warrant more investigation. School buildings look innocent from outside and yet students report feeling unsafe at school.

Regarding the effect of other school-level factors (*teacher-attributes* and *instructional-quality*), though their effects were not strongly predictive of school-mean achievement, they contributed in explaining the overall school effect. When the effect of *home-possessions* was modeled as a function of *teaching experience* (one of the teacher-attribute variables), the results indicated that in schools that had more experienced teachers (top 25% of experienced teachers), the variable *home-possessions* had more of an impact on achievement that in schools that had less experienced teachers (bottom 25%). The effect of this interaction is visually presented in Figure 2.

(Take in Figure 2 about here- interaction)

Figure 2 shows the effect of *home possessions* on *student achievement* across select values of teaching experience (top 25%, which have about 9 years of teaching in the case of Botswana, and bottom 25%, which have about 3 years of teaching experience). Besides the influence of *teaching experience*, the variable *home-possessions* had a constant effect on student achievement across schools. None of the aggregated variables seemed to explain away the importance of teaching experience as a moderating variable of the influence of home-possessions across schools on school mean achievement. Teaching experience need the attention of policy makers.

The other two cross level interaction effects (without necessarily representing them virtually), were between the level-1 variable *self-confidence* in mathematics and the school-level variable *degree of passive learning* (a measure of the Instructional-quality factor). Thus, the regression of *mathematics achievement* on *self-confidence* varied across the range of values of the *degree of passive learning*. The effect of *self-confidence* on achievement in mathematics was positive across schools over the range of values of the *degree of passive learning* (high values in the variable *degree of passive learning* were associated with school practices where there was less passive work) and the results indicated that the more self-confident students gained more with less passive teaching. The third interaction effect was between home-possessions and the number of books. The results indicated that the effect of the *number of books at home* on student achievement was stronger in some classrooms than in others. Specifically, the correlation between the *number of books* and *math achievement* was stronger in schools with high *mean home possessions*. This study could not fully explain why the effect of the variable *number of books* on achievement differed from school to school even after accounting for school-level variables. One possible explanation could have to do with the purpose that the books serve in the home (decoration versus information). It used to be in fashion in Botswana for households to own a set of the Britannica encyclopedia for prestigious purposes; some of these were hardly ever opened.

Conclusion and Implications

This study supported the notion that it is not because of the individual factors in isolation that the observed school effect is what it is, but the total effect of the indicator variables as a system. Six of the factors of the Rand model that were employed as a base for my hypothesis of school effectiveness were confirmed. Each of the factors contributed a significant indicator variable in explaining the education system in Botswana. Though the effect of the instructional quality factor was weak, its importance should not be underestimated. The weak effect could be due to the homogeneous classroom activities (similarity in teaching methods and class culture) as was the case with other teacher level characteristics. It could also be that the proxy variables used for this factor were not adequate. The original TIMSS nine items about classroom instruction were split into two proxies *passive* and *interactive* teaching because the effect of the composite variable made of all the nine items was too general for interpretation (and was not significant). It is hoped that with more analyses on the TIMSS teacher and classroom data, better understanding of the effect of the instructional quality factor will result.

In relation to the research question of which of the student and school level characteristics significantly explained students' mathematical achievement in Botswana and which of these variables were amenable to policy interventions: The findings in this study indicated that mathematics achievement was high among students who had *more books in their homes; spoke the language of the test (English) more frequently in their homes; possessed in their homes electricity, TV, Phone, flushing toilet, and a refrigerator; aspired for higher educational levels; valued mathematics; and felt confident to indulge in mathematics related activities.* The most significant student-level predictor was the *student educational aspiration.* Though most of the student level characteristics can not be manipulated through policy interventions, at the school level the results suggested that school mean achievement was high among schools with safe environment as perceived by the students and in schools where the student community (student body effect) aspired for higher educational attainment. That said, the results of this study suggested that the factors of *teacher attribute* and *instructional quality* served a moderating effect for student-level factors on achievement while the school quality factor had an overall effect on school mean achievement. The measures for these variables were mainly involved in cross-level interactions.

Implications for policy include making schools safe places for learning. There is need to investigate more about unsafe school conditions that students highlight so much as a problem and take action to address the problem. Another issue involves teacher retention. It does not augur well for the country to have the most inexperienced teachers internationally, especially given the benefits of long years of teaching on students performance. This other point concerns teacher education: There is need to foster positive attitudes among students in their learning for them to aspire more for higher education.

References

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14*, 219-234.

Allison, P.D. (2002). Missing data. Thousand Oaks, CA: Sage.

Broeck, A. Van Den., Opdenakker, M., & Damme, J. V. (2005). The Effects of Student Characteristics on Mathematics Achievement in Fleemish TIMSS 1999 data. *Educational Research and Evaluation, 11*(2), 107-121

Brookover, W. B., Schweitzer, J. H., Schneider, M. J., Beady, C. H., Flood, K. P., & Wisenbaker, J. M. (1978). Elementary school social climate and school achievement. *American Educational Research Journal, 15*(2), 301-318.

Bryk, A. S., & Thum, Y. M. (1989). The effects of high school organization on dropping out: An exploratory investigation. *American Educational Research Journal, 26*(3), 353-383.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. *Review of Research in Education, 8*, 158-233.

Campbell, J. R.., Hombo, C. M.; & Mazzeo, J. (2001). NAEP 1999 Trends in academic process: Three decades of student performance. *Education Statistics Quarterly 2*(4). Retrieved August 2, 2006, from http://nces.ed.gov/programs/quarterly/Vol_2/2_4/e_section2.asp#sec5.

Cele, C. I. (n.d.). The quest for fairness and equity in educational assessment in relation to school location. Retrieved July 20, 2006, from http://curriculum.wcape.school.za/resource_files/22152739_13.doc

George, R., & Kaplan, D. (1998). A structural model of parent and teacher influence on the science attitudes of eighth graders: Evidence from NELS: 88. *Science Education, 82*. 93-109.

George, R., & Kaplan, D. (1998). Evaluating latent variable growth models through ex post simulation. *Journal of Educational and Behavioral Statistics, 23*, 216-235.

Gibson, N. M., & Olejnik, S. (2003). Treatment of missing data at the second level of hierarchical linear models. *Educational and Psychological Measurement, 63*(2), 204 - 238.

Glasman, S. N., & Biniaminov, I. (1981). Input-output analyses of schools. *Review of Educational Research, 51*(4), 509-539.

Goldenberg, P. E. (2000). Thinking (and talking) about technology in math classrooms. In National Science Foundation (Ed.), *Issues in Mathematics Education*: Education Development Centre, Inc.

Gronna, S. S., & Chin-Chance, S. A. (1999). *Effects of school safety and school characteristics on grade 8 achievement: A multilevel analysis.* Paper presented at the Annual meeting of the American Educational Research Association

Hiebert, J., Stigler, J. W., Jacobs, K. J., Givvin, B. K., Garnier, H., Smith, M. (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 video study. *Educational Evaluation and Policy Analysis, 27*(2), 11-132.

Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam: TT-Publikaties.

Hox, J. (1998). Multilevel modeling: When and why. In I. Balderjahn, R. Mathar & M. Schader (Eds.), *Classification, data analysis, and data highways* (pp. 401-408). New York: Springer Verlag.

Kaplan, D., & Elliott, R. P. (1997). A model-based approach to validating education indicators using multilevel structural equation modeling. *Journal of Educational and Behavioral Statistics, 22*(3), 323-347.

Kaplan, D., & George, R. (1998). A structural model of parent and teacher influences on science attitudes of eighth graders: Evidence from NELS: 88. *Science Education, 82*(1), 93-110.

Kaplan, D., & Kreisman, M. B. (2000). On the validation of indicators of mathematics education using TIMSS: An application of multilevel covariance structure modeling. *International Journal of Educational Policy, Research, and Practice, 1*(2), 217-242.

Koller, O., Baumert, J., Clausen, M., & Hosenfeld, I. (1999). Predicting mathematics achievement of eighth grade students in Germany: An application of parts of the model of educational productivity to the TIMSS data. *Educational Research and Evaluation, 5*(2), 180-194.

Koretz, D., McCaffrey, D., & Sullivan, T. (2001). Using TIMSS to analyze correlates of performance variation in mathematics. Retrieved December, 12, 2007 from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=200105

Mandeville, G. K., & Liu, Q. (1997). The effect of teacher certification and task level on mathematics achievement. *Teaching and Teacher Education, 13*(4), 397-407.

Ma, X., & Wilkins, J. L. M. (2003). Modeling change in student attitude toward and beliefs about mathematics. *The Journal of Educational Research, 97*(1), 52-63.

Ministry of Education, (2005). *Report on the trends in international mathematics and science study (TIMSS) project for 2003*. Gaborone, Botswana: Government publishers

Mullis, V. S. I., Martin, O. M., Gonzalez, J. E., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report* MA: TIMSS & PIRLS International Study Center.

Oakes, J. (1986). *Educational indicators: A guide for policymakers*. Wisconsin-Madison, USA: Center for Policy Research in Education.

Peng, C. J., Harwell, M., Liou, S., & Ehman, L., H. (2006). Advances in missing data methods and implications for educational research. In C. Sawilosky (Ed.), *Real data* (pp. 31-78). Greenwich, CT: Information Age.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Application and data analysis methods* (2nd Ed.). CA USA: Sage Publications, Ltd.

Raudenbush, S. W., & Bryk, S. A. (1988-1989). Methodological advances in analyzing the effects of schools and classrooms on student learning. *Review of Research in Education*, *15*, 423-475.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R., & Toit, M. (2004). *HLM6 Hierarchical linear & nonlinear modeling: Application and data analysis methods.* . IL, U.S.A.: SSI Scientific software international, Inc.

Raudenbush, S. W., Bryk, A. S. & Congdon, R. (2005). HLM Software version 6.03 for windows

Rowan, B., Raudenbush, W. S., & Kang, S.J. (1991). Organizational design in high schools: A multilevel analysis. *American Journal of Education, 99*(2), 238-266.

Schafer, J. L., & Graham, W. J. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147 - 177.

Schulz, W. (2005). Mathematics Self-Efficacy and Student Expectations: Results from Pisa 2003. Paper presented at the annual meeting of the American Education Research Association, Montreal, 11-15 April.

Shavelson, J. R., McDonnell, M. L., & Oakes, J. (1989). *Indicators for monitoring mathematics and science education*.CA: The RAND Corporation.

Stevenson, H. W., Lee, S. Y., & Stigler, J. (1986). Mathematics achievement of Chinese, Japanese, and American children. *Science, 231*, 693-699.

Tabachnick, B., & Fidell, L. (2001). Using multivariate statistics (4th Ed.). MA: Allyn & Bacon.

Thomas & Heck. (2001). Analysis of large-scale secondary data in higher education research: Potential perils associated with complex sampling designs. *Research in Higher Education 42*(5), 517 - 540.

TIMSS 2003 databases. Retrieved May 30, 2004, from http://timss.org.

Wilkinson, L., & Task Force on Statistical Inference APA Board of Scientific Affairs. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*(8), 594-604.

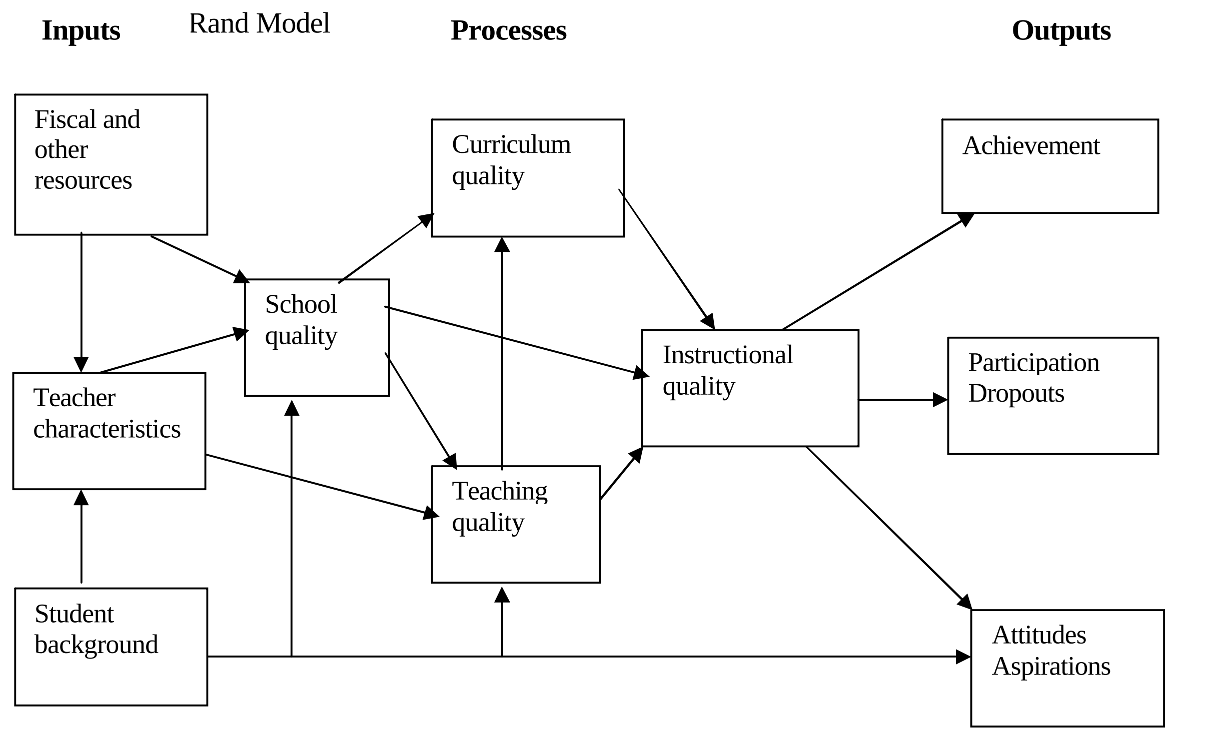
Walberg, H. J. (1981). A psychological theory of educational productivity. In F. H. Farley & N. Gordon (Eds.), *Psychology and Education* (pp. 81-110). Chicago, IL: National Society for the Study of Education.

Yang, Y. (2003). Dimensions of socio-economic status and their relationship to mathematics and science achievement at individual and collective levels. *Scandinavian Journal of Educational Research, 47*(1), 21-41.

Table 1. Factors, indicator variables, their corresponding metrics, descriptive statistics and HLM results for factors

Constructs (Factors)	Indicator variables	Questionnaire Item	**Metric**	**Mean**	**SD**	**a**	**HLMs for individual factors**		
							ß	*t*	*p*
Student Background	Intercept						367.6		
	Mother's education	Highest education level reached by mother	1=finished university to 4=primary or never [Recoded]	2.73	1.7	-	2.66	3.29	.01
Variance explained 10.2%	Books home	Number of books	1(200<) to 5 (<10) [Recoded]	1.84	1.06	-	5.09	3.9	.001
	Language home	Frequency of speaking English at home	1=always to 4=Never -TIMSS	2.02	.44	-	12.7	4.64	.001
	Home possessions	Electricity, TV, Refrigerator, Telephone (5)	Yes = 1 No = 2	1.66	.87	.826	7.05	4.38	.001
Attitudes	Intercept								
	Self Confidence	Take more math, usually do well (6)	1 agree a lot to 4=disagree a lot	2.65	.59	.564	14.32	7.2	.001
Variance explained 19.5%	Aspiration	How high in school student expects to go	1=finished university to 3=finish Secondary [Recoded]	3.14	1.69	-	12.7	17.47	.001
	Value math or utility of math	Need math for university/for other subjects/for daily life (5)	1 agree a lot to 4=disagree a lot	2.81	.44	.642	20.6	9.21	.001
Teacher attributes	Intercept						367		
	Experience	Number of years teaching	Continuous variable starting at 0	6.24	4.65	-	1.44	1.87	.064
Variance explained 16%	Beliefs	Different ways to solve math/sets of rules (5)	1 agree a lot to 4=disagree a lot	1.68	.44	.518	.201	.04	.967
	Qualification	Teachers education level	1=beyond university to 4 Finished Senior Sec [Recoded]	2.05	.35	-	15.69	1.76	.079
School Quality	Intercept						366.7		
	Locality	School locality (rural/semi-/urban)	1 = urban to 3 = rural [recoded]	1.64	.77	-	10.62	3.56	.001
Variance explained 28.7%	Climate (student)	Was hit/made fun of/ theft (4)	Yes = 1 No = 2 [recoded]	1.83	.62	.320	12.65	3.32	.002
	Climate (teachers)	School safety/teacher satisfaction (TIMSS)	1 = high to 3 = low - TIMSS	1.35	.54	.744	8.29	1.57	.120
Instructional Quality	Intercept						367.5		
Variance explained 10.2%	Interactive Learning	Explain reasoning/relate daily lives/decide own methods (6)	1= every day to 4 = Never	2.3	.48	.646	-3.03	-.48	.629
	Passive Learning	Practice numerical computations/fractions(3)	1= every day to 4 = Never	2.31	.47	.413	8.45	1.57	.120
Mathematics Achievement	Plausible Values	Test items	Dependent Variable	366	2.6				

Note. The first column gives the factors and the amount of variance explained by its measures when entered into an empty model,
Numbers in parenthesis in column 3 indicate the number of items used for the respective index variable;
a is a Reliability measure for the composite variables (Cronbach's alpha)
The last three columns depicts the results for separate HLM for the five factors *ß is the coefficient, then t-test and the p-value*



Note. The arrows indicate the direction of effect

(Oakes, 1986; Shavelson, McDonnell, & Oakes, 1989)

Figure 2. Interaction between *home possessions* and *teaching experience*

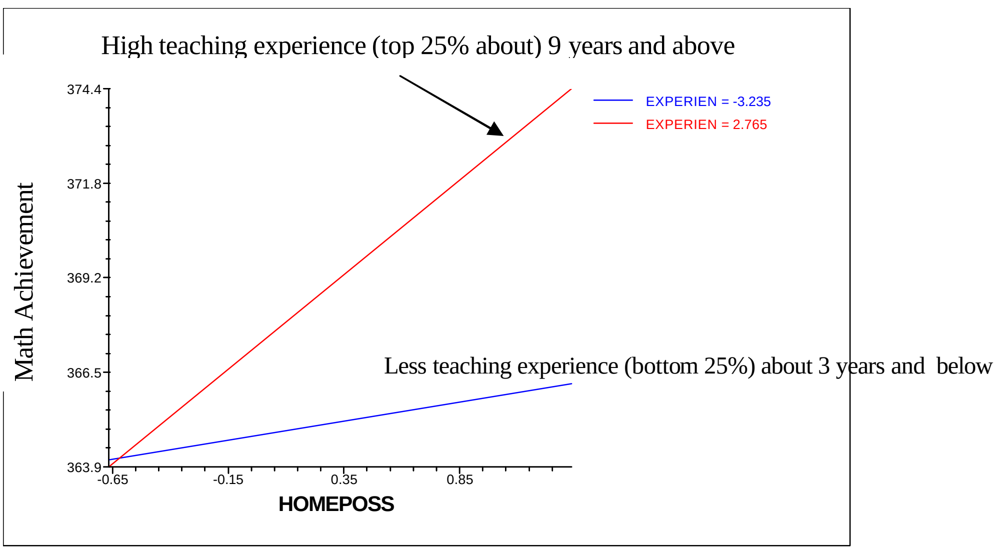


Figure 2. Interaction between home-possessions and teaching experience