The implications of a causal relationship between English language proficiency and academic performance

Yoko Miura

University of Cincinnati

Paper Presented at

The 21st Annual Convention of the University Council for Educational Administration

Alexandria, VA

November 17, 2007
The implications of a causal relationship between English language proficiency and academic performance

Abstract

This study investigated interactive or causal relationships among seven selected student variables (English language proficiency, gender, native language, socio-economic status, ethnicity, years of education in schools, and migrant status), two school variables (mobility and poverty) and academic achievement (test scores in reading, writing, and mathematics) using the Structural Equation Model (SEM) for the Limited English Proficient (LEP) students. The full SEM model consisted of two parts: the path-analytic model for the confirmatory factor analysis and the casual structure model. Findings indicate that there were causal and correlational relationships among variables chosen for this study. Overall, still a large portion of variance was explained to predict academic achievement of LEP students. The total effect of student background (indirect effect) and English language proficiency (direct effect) on academic achievement seemed to be about 87% of academic achievement for the LEP students. Although this study confirmed two hypothetical constructs to predict and explain academic achievement, overall results clearly indicate that the model may not be sufficient; therefore, the constructs could be questionable possibly by lacking significant amount of predictor variables. Student demographics were not sufficient predictors of LEP students’ English language attainment. A large portion of variances remained unexplained to predict English language proficiency of LEP students manifested in the English Language Development Assessment (ELDA) scores. The SEM results failed to prove the causality or associations between Whiteness and higher academic achievement. Generally speaking, the variable native language was predicted well by its predictors. The study also suggested a need for revising the term determining LEP status.

Statement of the Problem

The purpose of this study was to investigate interactive or causal relationships among seven selected student variables (English language proficiency, gender, native language, socio-economic status, ethnicity, years of education in schools, and migrant status), two school variables (mobility and poverty) and academic achievement (test scores in reading, writing, and mathematics among Limited English Proficient (LEP) students.

Many states report the achievement gap between LEP students and their non-LEP peers. Researchers have argued that high-stakes tests for LEP students may indeed be English language tests, and pointed out difficulties in making distinctions between language proficiency and academic competence that are assessed in high-stakes tests (Abedi, 2004; August & Hakuta, 1997, 1998; Baker, 1996; Coltrane, 2002; Crawford, 1999; Gonzalez, R.D. & Melis, 2000b; Hakuta & Mostafapour, 1996). Previous studies in Texas, California, and other states indicate that LEP students demonstrate lower academic achievement compared to their non-LEP peers and further suggest possible predictor variables (Coltrane, 2002; Legislative Analyst’s Office, 2005: Texas Education
Agency 1998; 2002). Ethnicity, native language, gender, socio-economic status, years-in-school, and migrant status are suggested to have high correlation with academic achievement of LEP students. Other studies point out that migrant status, and years in formal education in U.S. schools are correlated to socio-economic status, native language, ethnicity, and mobility (August & Hakuta, 1997; Baker, 1998; Coltrane, 2002; Goertz, Duffy, & Floch, 2001; Hakuta, Buttler & Witt, 2000; Holman, 1993; McLeod, 2000). Furthermore, campus mobility rate and campus poverty level may impact students’ performance, independent of their own mobility and socio-economic status (McLeod, 2000; Texas Education Agency 1998; 2002).

Research Questions

Research Question 1. Are there any inter-relationships between student variables (English language proficiency, gender, native language, socioeconomic status, ethnicity, migrant status, and longevity) and academic achievement (test scores in reading, writing, and mathematics) for the 4th grade and the 6th grade LEP students?

Research Question 2. Are there any inter-relationships among student variables (English language proficiency, gender, native language, socioeconomic status, ethnicity, migrant status, and longevity), school variables (campus mobility and campus poverty), and academic achievement (test scores in reading, writing, and mathematics) for the 4th grade and the 6th grade LEP students?

Research Hypotheses

Research Hypothesis 1.

H₀: There are no interactive or causal relationships between student variables (English language proficiency, gender, native language, socio-economic status, ethnicity, migrant status, and longevity) and academic achievement (test scores in reading, writing, and mathematics) in the state of Ohio.

Research Hypothesis 2.

H₀: There are no interactive or causal relationships among student variables (English language proficiency, gender, native language, socio-economic status, ethnicity, migrant status, and longevity), school variables (campus mobility and campus poverty), and academic achievement (test scores in reading, writing, and mathematics) in the state of Ohio.

Research Design & Methodology

The full SEM model consisted of two parts: the path-analytic model for the confirmatory factor analysis and the casual structure model. First, statistical assumptions of the SEM were checked before the construct models were created. The causal and interrelated relationships among variables exogenous and endogenous variables were determined with a priori hypothesis in the path analysis, namely, native language, gender, socio-economic status, migrant status, and longevity can predict English language proficiency, as manifested in the English Language Development Assessment (ELDA) scales. All of those individual variables along with race, and campus mobility and campus poverty can predict individual scaled scores in reading, writing, and mathematics.
on the Ohio Proficiency Test (OPT) and the Ohio Achievement Test (OAT) administered in March 2005.

The theoretical model (the path diagram among variables) was created using AMOS graphic to indicate directional and relationships in the diagram. The modification indices were used to improve the model fitness. The measurements of the model fitness were: the chi statistics to degree of freedom ratio, the Root Mean Square Error of Approximation (RMSEA), Adjusted Goodness of Fit Index (AGFI), Comparative Fit Index (CFI), Parsimony Comparative Fit Index (PCFI), and Akaike Information Criterion (AIC). The theoretical models were tested for the model fitness to obtain the final model for the confirmatory factor analysis of the paths and the validity of the model, then examined the full model.

**Statistical Model**

The Structural Equation Model (SEM) is a multivariate statistical technique. There are four steps in SEM: (1) the model specification; (2) the model identification; (3) the test of model fitness; and (4) the model modification. First, for the model specification, constructs were identified as either being exogenous or endogenous. For each construct, variables to indicate the measurement were determined. Then, to demonstrate the relationships easily, both exogenous and endogenous variables were portrayed in a path diagram using arrows to indicate correlational or directional relationship. The choice of which parameters are fixed and which are free (correlation versus covariance) should be consistent with the researcher’s a priori hypothesis. A theory-based model produces its own unique covariance matrix (Statistics Solutions, 2006). Second, the theoretical model was identified after the specifications of interactive paths between the demographic variables (English language proficiency, gender, native language, socio-economic status, ethnicity, migrant status, and years in schools) and test scores (Research Question 1), and the demographic variables, school variables (campus mobility and campus poverty), and test scores (Research Question 2). Third, the model estimation was conducted based on the methods that fit the sample size and the distribution of the data. The model modifications were performed based on the modification indices for assessing the measurement model validity (Hair, et al., 2005). The measurement model was then re-specified, assigning indicator variables to the constructs they should represent. Finally, the structural model validity was assessed using model goodness of fit indices.

**Data Preparation for the SEM**

The variables SES, gender, migrant, and migrant status were dichotomized data. All 4 ELDA scores were scaled according to the English Language proficiency level of the students. The variable longevity was also a scaled data indicating years spent in U.S schools. The variable race was originally classified into 6 categories, American Indian or Alaskan Native, Asian or Pacific Islander, Black (non-Hispanic), Hispanic, White (non-Hispanic), and multiracial. They were dichotomized into White and non-White. Also, the variable Native Language originally came with 23 codes including English as default, but they were dichotomized into Spanish and all others so that the data could be interpretable. All interval ratio data (campus mobility, campus poverty, three achievement scores in
reading, writing and mathematics) were standardized for the analysis.

Statistical Assumptions

The SEM is a multivariate analysis technique. The complexity of this technique requires meeting the statistical assumptions to increase the sensitivity. The statistical assumptions were tested prior to conducting the analyses to test the above two hypotheses. The statistical assumptions of the SEM are: (1) a large sample size; (2) an adequate amount and a pattern of the missing data; (3) univariate and multivariate normality and absence of outliers; (4) absence of multicollinearity and singularity; and (5) linearity between independent variables and dependent variables indicated by the residuals.

The Maximum Likelihood Estimation (MLE) method was chosen for this study due to its sustainability to statistical biases, although it is critical for this method to perform correct model specification because the MLE is a full estimation method (B. M. Byrne, 2001; R. B. Kline, 2005; R. H. Hoyle, 1995). The recommended sample size for the SEM is usually 200 or more, but the sensitivity of the Maximum Likelihood Estimation (MLE) is increased when the sample size is 400 or more (Tabachinick & Fidell, 2000). A larger sample size is less affected by problems that can occur by the missing data and in the data cleaning process. From the total number of 2,544 cases in the fourth grade and 1,985 cases in the sixth grade, this study examined 1,354 cases for the fourth grade and 1,168 cases for the sixth grade for the SEM. The selection process of the cases was not random, but simply based on the availability of the matching data provided by the Ohio state Department of Education as of June 2006. Due to the large data sample, the sensitivity to the effects of the missing data was still minimized. For the data preparation, the missing data were cleaned list-wise based on native language and English Language Development Assessment (ELDA) elements. There were less than 5% missing data after the list-wise deletion process. There were 5 missing data for math, 1 missing datum for reading, and 3 missing data for writing scores for the fourth grade. For the sixth grade, there were 3 missing data for math, 3 missing data for reading, and 43 missing data for writing scores. The requirement of Missing Completely At Random (MCAR) was fulfilled for all variables. The missing data were deleted list-wise, then substituted for the mean for math, reading, and writing scores (Hair, et al, 2005). Because the missing data for the campus mobility rate and the campus poverty rate were concentrated in a few school buildings, the pattern of the missing data for those two variables was of concern. For those, the mean substitution was adopted as a remedy instead of list-wise deletion.

The sample size also affects (1) the multivariate distribution of the data, (2) the model complexity, (3) the amount and the pattern of the missing data, and (4) the amount of average error variance among the reflective indicators. Both the univariate and the multivariate normality were examined. First, the skewness and the kurtosis were in the acceptable range, -2.58 < Z skewness < 2.58 (p < .01) except for the variable “mobility” in the fourth grade and in the sixth grade. The histograms for interval-ratio variables indicated the normal shape of the data distribution for all variables; however, the campus mobility rate and the migrant status indicated that the data were positively skewed for both the fourth and the sixth grades. The outliers were also checked, but not removed until the examination of the multivariate normality. Extreme values and linear
relationships in the data were checked using Multiple Regression Analysis. The model summary for the Analysis of Variation (ANOVA) indicated all independent variables were significantly associated with the three dependent variables (math, reading, and writing scores) at \( p < .01 \) level. The scatter plots of the standardized residuals among all variables were examined for the linear relationships between the dependent variables and the independent variables. The scatter plots indicated that the residuals were small and centered around the zero. This met the criterion for the assumption of the homoscedasticity. Finally, outliers were identified using Mahalanobis distance divided by the number of variables for the approximate t-value using \( D^2/df < 4 \) as a cut off point.

There was one multivariate outlier for the fourth grade data, and there was no multivariate outlier for the sixth grade data. The outlier was removed before the SEM analyses. There were linear relationships between the dependent variables, academic scores (reading, writing, and mathematics tests) and independent variables, all other variables (SES, race, gender, migrant, native language, longevity, 4 ELDA scores, mobility, and campus poverty). For the multicollinearity and the singularity, correlations were examined among all variables was conducted. Generally, using the cut off point of .8, no multicollinearity problems were found. When a multicollinearity increases, “it complicates the interpretation of relationships because it is more difficult to ascertain the effect of any single construct owing to their interrelationships” (Hair, Black, Babin, Anderson, and Tatham, 2005, p.709). From all tests used for checking the statistical assumptions, it is fair to say that all statistical assumptions for the SEM were met.

The Confirmatory Factor Analysis for the Path-Analytic Models

The first step of the SEM for this study was to identify theoretical and hypothetical constructs for the observed variables. Three steps were considered for the process of the model specification in this study. First, the confirmatory factor analyses for the path-analytic models were examined with three possible models with different numbers of constructs and relationships among them without dependent variables. The second step was to specify the paths among the independent variables in order to achieve the best model fitness. Finally, the full causal structure models were examined with dependent variables, achievement in reading, writing, and mathematics.

For the first step, three basic paths models were considered to address the research hypothesis 1 considering (1) number of constructs and (2) relationships between constructs and observed variables. To test the validity of the constructs and their relationships, reasonable paths of covariance and the parameter weights were added to each model based on the theoretical assumptions and information obtained in the process of preliminary analysis for examining the statistical assumptions for the SEM. For all three models, the parameter weights for student background were set on the variable longevity for the prediction of academic achievement. For the two construct path-analytic models, which were models 2 and 3, the variable listening also had the parameter weight for the construct English proficiency. For all 3 models, the paths were added between longevity and migrant status, race and migrant status, SES and migrant status, and migrant status and native language. Estimates were calculated using the AMOS 6 program for regression weights, standardized weights, parameter estimates, assessment of multivariate abnormality, and modification indices.

Figures 1, 2 and 3 illustrate the paths and the constructs. The first paths model
consisted of one hypothetical construct, “student background” for manifesting the six demographic variables and all four ELDA scales together. It was based on the hypothesis that student background consists of all 10 variables. Figure 1 illustrates the initial path-analytic model for the first model for the fourth grade.

Figure 1. Model 1: Initial 4th grade path-analytic model with one construct, student background.

The second model consisted of two hypothetical constructs. One was “student background” to represent the six demographic variables, longevity, SES, race, gender, migrant, and native language. The other was “English proficiency” to represent the four ELDA tests of ELA elements in listening, speaking, reading, and writing. In this model, two constructs were correlated, in other words, this model assumed a correlation
relationship between two constructs and observed variables which manifested the two hypothetical constructs. Figure 2 illustrates the initial path-analytic model 2 with two constructs and a correlation between the two constructs for the fourth grade.

The third model consisted of two constructs. One was the construct “student background” to represent six demographic variables, longevity, SES, race, gender, migrant, and native language. The other was the construct “English proficiency” to represent four ELDA tests of ELA elements in listening, speaking, reading, and writing. In this model, two constructs had a causal relationship, that was, student background influences English proficiency. Figure 3 illustrates the path-analytic model 3 with two constructs and a causal, thus directional relationship between the constructs for the fourth grade.
Table 1 (page 10) lists the results of all three models.

For the model assessment, two aspects were discussed, (1) parameter estimates and (2) model as a whole. There were three aspects of concern in examining individual parameter estimates: (1) feasibility of parameter estimates, (2) appropriateness of standard errors, and (3) the statistical significance of the parameter estimates (Byrne, 2000). For the model assessment as a whole concerned several indices of the goodness-of-fit statistics. First of all, the initial path-analytic model 1 was assessed as not an admissible solution by the AMOS 6.0 software. This means that either the model 1 was wrong, or the input matrix of the model 1 lacked sufficient information. In other words, the model 1 was assessed as mis-specified, or the model with one construct “student background” which was inclusive of all demographic variables and English language proficiency did not fit the data. Second, for the assessment of the model fitness for models 2 and 3, the $\chi^2$ statistics must not be significant ($p > .05$), but the results for all 3 models did not indicate it ($p < .001$).

Figure 3. Model 3: Initial 4th grade path-analytic model with two constructs, student background and English proficiency and their causal relationship.
Table 1: *Fit Indices for the 4th Grade Basic Path-Analytic Models*

<table>
<thead>
<tr>
<th>Models</th>
<th>$\chi^2$</th>
<th>RMSEA</th>
<th>AGFI</th>
<th>CFI</th>
<th>PCFI</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. One construct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Two Constructs</td>
<td>155.26***</td>
<td>.057</td>
<td>.956</td>
<td>.961</td>
<td>.620</td>
<td>207.26</td>
</tr>
<tr>
<td>(correlational)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Two Constructs</td>
<td>141.65***</td>
<td>.055</td>
<td>.959</td>
<td>.965</td>
<td>.601</td>
<td>195.65</td>
</tr>
<tr>
<td>(causal)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. * $p < .05$, ** $p < .01$, *** $p < .001$. RMSEA, Root Mean Square Error of Approximation; AGFI, Adjusted Goodness of Fit Index; CFI, Comparative Fit Index; PCFI, Parsimony Comparative Fit Index; and AIC, Akaike Information Criterion. Byrne (2000) stated that more recently Bentler recommended PCFI over CFI, and CFI over NFI (Bentler and Bonnet’s Normed Fit Index, 1980). a. AIC is a most commonly reported predictive index, which is to be compared with the value of the saturated model (AIC = 110). Brynes (2000) and Kline (2005) argued that CAIC (Consistent AIC) had been reported to be more appropriate because of its adjustment to sample size, but this study had a large sample size. For the same reason, the Bayes Information Criterion (BIC), which penalizes for the model complexity, was not considered in this study.*

However, the AMOS program still assessed the minimum was achieved, which means the solution was acceptable, because the $\chi^2$ statistics is almost always significant for a very large sample size study. This study is considered to be a very large sample size study. In other words, the $\chi^2$ statistics alone does not provide sufficient interpretation in the SEM analyses. For the model 2, the number of the sample moments were 55, and the $\chi^2$ statistics was 155.26 ($df = 29$, $p < .001$). This means that the model was over-identified. In the SEM, in order to reject the null hypothesis, we must create the model which is over-identified, otherwise, there would be no null hypothesis to reject, thus no need of the scientific inquiry (Byrne, 2000, Klein, 2005). For this model, the overall model fitness indicator RMSEA was .057. The model is generally considered to be in an acceptable range when $0.05 < \text{RMSEA} < 0.08$, and not acceptable when $\text{RMSEA} > 0.10$. For a good fitness, the model must achieve RMSEA < .05, which will provide PCLOSE, which would give a test of close fit with the probability value. In the path-analytic model 2, AGFI, an absolute fit index, was estimated as .956 and an incremental fit index, CFI = .961 showed that this model indicated a very good fitness for baseline comparisons. Next, for the parsimony-adjusted measurement, PCFI a parsimony adjusted CFI, sensitive to model size was .620, which was in an acceptable range as “Mulaik et al, suggested that non-significant $\chi^2$ statistics and goodness-of-fit indexes in the .90s, accompanied by parsimonious-fit indices in the .50s, are not unexpected” (Byrne, 2000, p.82). Parsimony-
adjusted measurements could provide indicators for a better and more parsimonious model fitness to be compared with this model. A predictive index, AIC (207.256) could be closer to the saturated model (AIC = 110).

For the model 3, the $\chi^2$ statistic was 141.65 ($df = 28, p < .001$), and the overall model fitness indicator RMSEA was .055. The model was considered to be in an acceptable range to proceed with further analyses. In this model, other fit indices were .959 for AGFI and .965 for CFI both showed that this model indicated the good fitness, lightly better than that of the model 2. AIC was 195.65, like the model 2, it could be closer to the saturated model (AIC = 110). PCFI of .601 indicated that the model 2 was slightly better than this model 3, but the value of the saturated model was .0. There were not dramatic differences between the path-analytic models 2 and 3. Therefore, both models remain to further specify the most interpretable models for the data.

The model specification. The modification indices of the path-analytic models 2 and 3 suggested possible improvements for the model. This part was rather exploratory. The paths were added according to the modification indices between longevity and SES, race and writing, and ELDA’s reading scale was added to manifest both student background and English proficiency for the path-analytic model 2 assumed a correlational relationship between two constructs, student background and English proficiency. For the path-analytic model 3, only a path between longevity and SES was added. Figures 4 and 5 show the final path-analytic models with estimates.
Figure 4. The final 4th grade result of the path-analytic model 2: Two constructs, student background and English proficiency and their correlational relationship.
Figure 5. The final 4th grade result of the path-analytic model 3: Two constructs, student background and English proficiency and their causal relationship.

Table 2 shows the summary of the model assessment. For the path-analytic model 2, the $\chi^2$ statistics was 102.10 ($df = 25, p < .001$), RMSEA was improved from .057 to .048. RMSEA < .05 gave the PCLOSE, the probability for obtaining RMSEA of .048 as 63.2%. Overall, the model improved; AGFI was from .956 to .959, CFI from .961 to .976, and AIC from 207.26 to 162.10. For the path-analytic model 3, the $\chi^2$ statistic was 101.98 ($df = 27, p < .001$), RMSEA was improved from .055 to .045. RMSEA < .05 gave the PCLOSE, the probability for obtaining RMSEA of .045 as 78.3%. Overall, the model improved; AGFI was from .959 to .970, CFI from .965 to .977, and AIC from 195.65 to 157.98.
Table 2

*Fit Indices for the Final 4th Grade Path-Analytic Model Comparison*

<table>
<thead>
<tr>
<th>Models</th>
<th>$\chi^2$</th>
<th>RMSEA</th>
<th>AGFI</th>
<th>CFI</th>
<th>PCFI</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Basic Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Two Constructs (correlational)</td>
<td>155.26***</td>
<td>.057</td>
<td>.956</td>
<td>.961</td>
<td>.620</td>
<td>207.26</td>
</tr>
<tr>
<td>3. Two Constructs (causal)</td>
<td>141.65***</td>
<td>.055</td>
<td>.959</td>
<td>.965</td>
<td>.601</td>
<td>195.65</td>
</tr>
<tr>
<td><strong>Final Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Two Constructs (correlational)</td>
<td>102.10***</td>
<td>.048</td>
<td>.968</td>
<td>.976</td>
<td>.542</td>
<td>162.10</td>
</tr>
<tr>
<td>3. Two Constructs (causal)</td>
<td>101.98***</td>
<td>.045</td>
<td>.970</td>
<td>.977</td>
<td>.586</td>
<td>157.98</td>
</tr>
</tbody>
</table>

*Note.* *p* < .05, **p* < .01, ***p* < .001. RMSEA, Root Mean Square Error of Approximation; AGFI, Adjusted Goodness of Fit Index; CFI, Comparative Fit Index; PCFI, Parsimony Comparative Fit Index; and AIC, Akaike Information Criterion.

Lastly, the interpretability of the models was reconsidered. Although there was a slight decline in PCFI in model 3, the penalty of the model complexity was clearly less on the model 3 than on the model 2. According to the modification indices, there could have been possible manifestation of the construct student background by ELDA speaking and ELDA listening as well. However, such a model is not interpretable. If all these paths would have been added, there would be no justification to dismiss the model 1, which only had one construct, student background. As indicated in Figure 8, the final model 3 clearly shows a simpler model, which was also more interpretive. In this model, there was a causal relationship between the construct student background and the construct English proficiency. RMSEA is desirable to be lower than .05, and the model 3 indicated lower RMSEA than the model 2, which gave the higher PCLOSE for the model 3 (78.3%) compared to the model 2 (63.2%). Therefore, the path-analytic model 3 was adopted for the full model causal structure analysis.
The Full Causal Structure Models

Figure 6 shows the full causal structure model for the fourth grade LEP students in Ohio.

Figure 6. The full causal structure model for the 4th grade LEP students in Ohio.
For the basic causal structure model, the $\chi^2$ statistic was 414.82 ($df = 56$, $p < .001$), and the overall model fitness indicator RMSEA was .069. The model was considered to be in an acceptable range to proceed with further analyses, but it needed improvement for obtaining a reliable PCLOSE statistic. In this model, other fit indices were .926 for AGFI and .946 for CFI. Both showed that this model indicated good fitness. AIC was 484.82; it could be closer to the saturated model (AIC = 182). PCFI was .679, supporting good fitness as well. As stated earlier, PCFI could be in .50s when other indices indicate $> .90$, and the $\chi^2$ statistic is non-significant. Again, in the SEM, “findings of well-fitting hypothesized models, where the $\chi^2$ value approximates the degree of freedom, have proven to be unrealistic in most SEM empirical research” (Bryne, B.M., 2000, p.81). The modification indices of the basic causal structure model suggested possible improvements for the model. This part was again, rather exploratory. A serious consideration was given to dropping the variable migrant status, which raised concerns about the data especially for the ELDA scales because of the regression weights estimation from the basic causal model. The probability of obtaining the critical ratio was $p > .05$, indicated as non-significant. If the result of the model with migrant status failed, this consideration would have been carried on, because the model with less complexity, which means, fewer paths, less variables, and less constructs are considered to be better if the tests of fitness are about the same. However, the research questions were interested all variables listed in the model and their interrelationships. Therefore at this point, the elimination of any variables was not considered. For the next step, modifications were made by adding the paths according to the modification indices between variables. The paths were made among endogenous variables; the OAT reading and writing and the OPT mathematics, between longevity and ELDA speaking, SES and ELDA speaking, native language and ELDA speaking, ELDA speaking and ELDA listening, and ELDA speaking and ELDA reading. Figure 7 shows the final causal structure model with estimates for the fourth grade LEP students in Ohio.
Figure 7. The final causal structure model with estimates of total effects (standardized regression weights) and squared multiple correlations for the 4th grade LEP students in Ohio.
The null hypothesis of the research question 1 was rejected. There were interactive or causal relationships among students variables (English language proficiency, gender, native language, socio-economic status, ethnicity, years of education in schools, and migrant status) and academic achievement (test scores in reading, writing, and mathematics) in the fourth grade and the sixth grade in the state of Ohio.

For the fourth grade LEP students in Ohio, most regression weights for student background in the prediction of race, SES, longevity, and native language were significantly different from zero at \( p < .05 \) level. Regression weights for student background in the prediction of gender were not significantly different from zero at \( p < .05 \) level. Gender had no significant correlations with any other variables. Regression weights for student background in the prediction of migrant status were not significantly different from zero at \( p < .05 \) level. Migrant status had weak but significant correlations with variables race (-.10), longevity (-.14), and native language (.27) \( p < .05 \). SES had weak but significant correlations with variables longevity (.23) and speaking scale of English proficiency (.07) at \( p < .05 \) level. The standardized total (direct and indirect) effect of English proficiency on academic achievement was .87. The predictors of academic achievement predicted 75.5\% of its variance. The standardized total effect of student background on English proficiency was .38. The predictors of English proficiency predicted only 14.1 \% of its variance. The predictors of native language predicted 53.0 \% of its variance. The speaking scale had mild correlation with the listening scale in English proficiency (.40), yet the predictors of the speaking scale and the listening scale predicted only 21.5 \% of the variance of the speaking scale and 51.6 \% of the variance of the listening scale. The listening scale also had a weak correlation with longevity (.29). The predictors of the reading scale of English proficiency predicted 84 \% of its variance. Among academic achievement, the predictors of mathematics predicted a lower portion of its variance (58.8 \%) compared to reading (79.3 \%) and writing (68.6 \%).

Table 3 shows the comparison of the results between path-analytic models and final full causal structure models for the fourth grade LEP students in Ohio.

<table>
<thead>
<tr>
<th>Models</th>
<th>( \chi^2 )</th>
<th>RMSEA</th>
<th>AGFI</th>
<th>CFI</th>
<th>PCFI</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Path-analytic models</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>141.65***</td>
<td>.055</td>
<td>.959</td>
<td>.965</td>
<td>.601</td>
<td>195.65</td>
</tr>
<tr>
<td>Final</td>
<td>101.98***</td>
<td>.045</td>
<td>.970</td>
<td>.977</td>
<td>.586</td>
<td>157.98</td>
</tr>
<tr>
<td>Final full causal structure models</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>414.82***</td>
<td>.069</td>
<td>.926</td>
<td>.946</td>
<td>.679</td>
<td>484.82*</td>
</tr>
<tr>
<td>Final</td>
<td>168.26***</td>
<td>.040</td>
<td>.968</td>
<td>.983</td>
<td>.668</td>
<td>244.26</td>
</tr>
</tbody>
</table>
Note. * $p < .05$, ** $p < .01$, *** $p < .001$. RMSEA, Root Mean Square Error of Approximation; AGFI, Adjusted Goodness of Fit Index; CFI, Comparative Fit Index; PCFI, Parsimony Comparative Fit Index; and AIC, Akaike Information Criterion. 
a. AIC value for the saturated model was 182.

For the final causal model, the $\chi^2$ statistic was 168.26 ($df = 53$, $p < .001$), RMSEA was improved from .069 to .040. RMSEA < .05 gave the PCLOSE, the probability for obtaining RMSEA of .040 as 99.2%. Overall, the model improved; AGFI was improved from .926 to .968, CFI was improved from .946 to .983, and AIC from 484.82 to 244.26. Compared to the path-analytic model, RMSEA was improved from .045 to .040, AGFI was from .970 to .968, and CFI from .977 to .983. Even PCFI, after parsimony adjusted and taking account the sample size, was better compared to the path-analytic model from .586 to .668.

Next, the sixth grade data were examined using the same casual structure. Figure 8 indicates the basic causal structure model obtained for the sixth grade LEP students in Ohio.
Figure 8. The basic 6th grade causal structure model for LEP students in Ohio.
In the basic causal structure model for the sixth grade, the $\chi^2$ statistic was 441.68 ($df = 56, p < .001$), and the overall model fitness indicator RMSEA was .078. The model was considered to be in an acceptable range to proceed with further analyses, but it needed improvement for obtaining a reliable PCLOSE statistics. In this model, other fit indices were .901 for AGFI and .937 for CFI. Both showed that this model indicated the acceptable fitness. AIC was 511.678; it could be much closer to the saturated model (AIC = 182). PCFI was .673, supported good fitness as well. The modification indices of the basic causal structure model suggested the model definitely needed improvements.

For the next step, modifications were made by adding the paths according to the modification indices between variables. The paths were made among endogenous variables; the OPT reading and writing, between gender and reading, ELDA speaking and ELDA listening. Also, the causal paths from SES to mathematics and native language to mathematics were added. Figure 9 shows the final causal structure model with estimates for the sixth grade LEP students in Ohio.
Figure 9. The final 6th grade causal structure model with estimates of total effects (standardized regression weights) and squared multiple correlations for LEP students in Ohio.
For the sixth grade LEP students in Ohio, most regression weights for student background in the prediction of race, SES, longevity, native language, and migrant status were significantly different from zero at \( p < .05 \) level. Regression weights for student background in the prediction of gender were not significantly different from zero at \( p < .05 \) level. Gender had no significant correlations with any other variables. Native language had weak but significant correlations with variables longevity (.19), and migrant status (.14) at \( p < .001 \). SES had weak but significant correlations with variables longevity (.20) at \( p < .001 \) level. The causal path from SES to mathematics was weak but significant (-.15), and the causal path from native language to mathematics was also weak but significant (.08). The standardized total (direct and indirect) effect of English proficiency on academic achievement was .87. The predictors of academic achievement predicted 75.2% of its variance. The standardized total (direct and indirect) effect of student background on English proficiency was .48. The predictors of English proficiency predicted only 22.9% of its variance, native language predicted 42.5% of its variance, the speaking scale and the listening scale predicted only 35.2% of the variance of the speaking scale, but 65.2% of the variance of the listening scale, 75.9% of the variance of the reading scale, and 74.7% of the variance of the writing scale. The listening scale had a moderate but significant correlation with the speaking scale (.41). Among academic achievement for the sixth grade LEP students in Ohio, the predictors of mathematics (58.7%) and writing (58.9%) predicted a lower portion of its variance compared to reading (90.3%). The direct paths from SES and native language improved the probability of obtaining RMSEA at the significant level (82% instead of 35%), but they did not improve the predicting portion for the mathematics achievement.

Table 4 shows the comparison between the basic model and the final causal model for the sixth grade.

Table 4

<table>
<thead>
<tr>
<th>Models</th>
<th>( \chi^2 )</th>
<th>RMSEA</th>
<th>AGFI</th>
<th>CFI</th>
<th>PCFI</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic</td>
<td>441.68***</td>
<td>.078</td>
<td>.901</td>
<td>.937</td>
<td>.673</td>
<td>511.68*</td>
</tr>
<tr>
<td>Final</td>
<td>179.19***</td>
<td>.046</td>
<td>.958</td>
<td>.979</td>
<td>.666</td>
<td>255.19</td>
</tr>
</tbody>
</table>

Note. * \( p < .05 \), ** \( p < .01 \), *** \( p < .001 \). RMSEA, Root Mean Square Error of Approximation; AGFI, Adjusted Goodness of Fit Index; CFI, Comparative Fit Index; PCFI, Parsimony Comparative Fit Index; and AIC, Akaike Information Criterion.

a. AIC value for the saturated model was 182.

In the final causal model for the sixth grade, the \( \chi^2 \) statistic was 179.19 (\( df = 53, p < .001 \)), RMSEA was improved from .078 to .046. RMSEA < .05 gave the PCLOSE, the probability for obtaining RMSEA of .040 as 82.0%. Overall, the model improved; AGFI
was improved from .901 to .958, CFI was improved from .937 to .979, and AIC was improved from 511.68 to 255.19. The concern was the value of PCFI, after parsimony adjusted and taking account of the sample size. It declined from .673 to .666, although it is not unexpected as discussed earlier.

From all of the results of the research hypothesis 1, it could be concluded that for both the fourth grade and the sixth grade LEP students’ data, there were causal relationships between the constructs “student background” and “English proficiency”, and “English proficiency” and “Academic achievement.” The results also confirmed that there were correlational relationships among observed variables, although the relationships slightly differ between the fourth grade LEP students’ data and the sixth grade LEP students’ data. In the fourth grade, the paths were much simpler than those of the sixth grade. Although the variable migrant status did not indicate statistical significance to manifest the construct student background, perhaps due to its homogeneity, it had significant effects to other observed variables such as race, SES, longevity and native language. Gender appeared to have little effects; however, there were significant relationships between gender and variables writing and mathematics in the construct academic achievement. On the other hand, in the sixth grade results, native language seemed to carry more weight in effects. Gender and migrant status appeared to be statistically insignificant as total effects on the construct student background, but had influence on other variables. Next, the school variables were added to the causal structure models to investigate research hypothesis 2.

Research hypothesis 2. Considering the results of research hypothesis 1 for the fourth grade and the sixth grade, the model with two additional school level variables, campus poverty and campus mobility was examined for both the fourth grade and the sixth grade accordingly. Figure 10 illustrates the hypothetical casual structure model for this hypothesis for the fourth grade LEP students.
Figure 10. The hypothetical causal structure model for the 4th grade LEP students in Ohio: Influence of student and school variables on English proficiency and academic achievement.
AMOS 6.0 program assessed this solution was not admissible for both the fourth grade and the sixth grade LEP students’ data. This indicated that some variance estimates were negative, and the model specification was wrong, or the sample size was too small. For this study, the sample size was quite large (> 400) for both grades. Therefore, it could be concluded that the effort to investigate any interrelationships, causal or correlational, among student, school, and academic achievement failed. The research hypothesis 3b could not be further examined.

Findings
Findings indicate that there were causal and correlational relationships among variables chose for this study. Overall, still a large portion of variance was explained to predict academic achievement of LEP students. The total effect of student background (indirect effect) and English language proficiency (direct effect) on academic achievement seemed to be about 87% of academic achievement for both the fourth grade LEP students and the sixth grade LEP students.

Although this study confirmed two hypothetical constructs to predict and explain academic achievement, overall results clearly indicate that the model may not be sufficient; therefore, the constructs could be questionable possibly by lacking significant amount of predictor variables. Student demographics were not sufficient predictors of LEP students’ English language attainment. A large portion of variances remained unexplained to predict English language proficiency of LEP students manifested in the ELDA scores.

Predictors of academic achievement for both grades seemed to explain about 75% of it variance, although the variance explained by student background on English proficiency. Only 14.1 % in the fourth grade and 22.9 % in the sixth grade English proficiency were predicted by student background. It suggests the existence and significance of other variables which were not included in this study’s model. The possible missing variables could be motivation and attitude, years and type of former education in the first language, school variables such as school climate, school leadership and teacher factors, or parental support and home, teen, school, or neighborhood culture.

Conclusions
There is no need to be concerned that LEP students, non-White students, or Hispanic students will lower learning standards (or something similar). Schools with LEP students had wide ranges of average mean test scores within schools just as all other schools. The performance of LEP students does not cause the school to lower the mean achievement scores in reading, writing, or mathematics. The SEM results failed to prove the causality or associations between the variables Whiteness and higher academic achievement, and the HLM results failed to prove the causality or associations between Hispanic students and lower achievement.

Generally speaking, native language was predicted well by its other student background variables (53 % of its variance in the fourth grade and 42.5 % of its variance in the sixth grade). Native language had weak but significant correlations with variables longevity and migrant status. The predictors of speaking scale and listening scale predicted only 35.2% of the variance of speaking scale, but 65.2 % of the variance of listening scale, 75.9 % of the variance of reading scale, and 74.7 % of the variance of
CAUSAL RELATIONSHIP

writing scale in the sixth grade data. Listening scale also had a moderate but significant correlation with speaking scale. There is a need for revising the definition of LEP status. The term LEP is limited to students who have spent fewer than 12 months in U.S. schools, but does not mean students who have lived in the U.S. for fewer than 12 months. Practically, living conditions and schooling of newly arrived immigrants differ from the norm of the U.S. school age children. Actual attendance in the U.S. schools may not be equal to the residency in the U.S. schools in many cases. There may be a discrepancy between students’ enrollment on the database and actual daily membership in schools especially among LEP students. On the other hand, instructional placement of the students is determined by English Language Assessment of the ELDA.

Future studies may be needed but not limited to the investigations of (1) large missing data, (2) students’ psychological variables, attitude and motivation, personalities and learning styles, (3) teen culture, home culture, parents and home environment, neighborhood culture (4) dichotomization of race and native languages, (5) longitudinal effects on the same cohort until graduation, (5) economic factors, such as school funding and socio-economic status, (6) instructional programs for LEP students, and (7) LEP students identified with special needs.
References


