Spaan Fellow Working Papers in Second or Foreign Language Assessment
Copyright © 2008
Volume 6: 29–55

English Language Institute
University of Michigan
www.lsa.umich.edu/eli/research/spaan

Investigating the Invariance of the ECPE Factor Structure across Different Proficiency Levels

Anja Römhild
University of Nebraska

ABSTRACT This study investigates the degree of factor differentiation in the factor structure of the Examination for the Certificate of Proficiency in English (ECPE) across two proficiency level groups. The study uses a split-half technique to obtain an independent measure of language proficiency from one ECPE test half and uses the other test half as the dependent variable. Analyses are based on item-level data from the multiple-choice section of the ECPE and involved multiple group confirmatory factor analysis to test the degree of measurement and structural invariance between the two proficiency groups. The results indicate that four of the five subtests exhibit partial measurement invariance across the low and high proficiency group. Only the cloze subtest showed full measurement invariance. Furthermore, none of the factor correlations and factor variances were invariant. The pattern of the factor correlations indicated weaker correlations among the factors in the low proficiency group and stronger correlations in the high proficiency group. This suggests that, for the particular examinees in this study, the structure of the test becomes increasingly more undifferentiated as language proficiency increases.

A major focus in language testing research has been the study of the factorial structure of language proficiency tests. Examining the factor structure of a test plays an important role in establishing validity evidence for the measurement of theoretical constructs such as language proficiency (Nunnally, 1978) and in evaluating whether a test exhibits equivalent measurement properties across different examinee populations (i.e., measurement invariance; Meredith, 1993). An important issue that arises from previous studies investigating the dimensionality of language tests is that the factor structure changes across examinee groups with different levels of language proficiency. Specifically, it has been found that the level of proficiency of the examinee group is related to the degree of factor differentiation exhibited by the test (indicated by a greater number of factors with higher factor loadings and lower factor correlations). However, not all research studies found the same relationship between proficiency level and factor differentiation. At least two studies observed that language proficiency is positively related to the degree of factor differentiation exhibited by the test (Swinton & Powers, 1980; Ginther & Stevens,1995). Other research studies, however, show a negative relationship with a decrease in factor differentiation as language proficiency increases (Hosley & Meredith, 1976; Oltman, Stricker, & Barrows, 1988; Kunnan, 1992). In addition, at least one study (Shin, 2005) concludes that there is insufficient evidence to suggest that factor differentiation occurs, even though this study found differences in the factor structure across different proficiency groups.

Another factor that has complicated the interpretation of existing research has been the lack of comparability of individual studies due to differences in the methodologies employed. In particular, differences in the methods of analysis and the type of input data chosen have made generalization of the research outcomes difficult. For example, while most of the studies employed factor-analytic methods, few used statistical tests of nested model comparisons to evaluate the statistical significance of observed differences in model parameter estimates (e.g., Ginther & Stevens, 1995; Shin, 2005). This method is preferable, however, because it evaluates model differences according to a specified statistical criterion and avoids the risk of capitalization on chance associated with exploratory methods. Several studies have also examined factor structures based on subscale scores rather than item-level data (e.g., Hosley & Meredith, 1976; Kunnan, 1992; Shin, 2005). This approach is problematic because it relies on the assumption that the measurement of the subscale scores is equivalent across groups.

Given the inconsistencies in the findings and design of past research studies, there continues to be a need to study the relationship between language proficiency and test functioning. In particular, it is of interest to investigate more closely under what circumstances systematic changes in the factor structure appear in relation to changing proficiency levels. In addition, there is a need for more rigorous description and analysis of the types of changes that are observed focusing not only on structural differences but also on differences in the measurement properties of the test. This should also involve the use of more rigorous statistical analysis tools that avoid the problem of capitalization on chance. Finally, this research has direct implications for the validity of score interpretations from language proficiency instruments. If the factor structure of a test varies across examinees as a function of their language proficiency level, then score comparisons are no longer meaningful because different kinds of information are gained from the test score for different groups of examinees. As a consequence, test developers may need to rethink the use of composite scores for multiple language skills and knowledge components.

Study Objectives

The purpose of this study is to examine the factorial structure of the Examination for the Certificate of Proficiency in English (ECPE) and to compare it across two proficiency level groups (low and high). The goal is to determine if the measurement properties and the structural relationships of the test differ across the two proficiency groups and to evaluate if these differences indicate systematic changes of either increasing or decreasing factor differentiation. The test instrument in this study is an English language proficiency exam intended to be used for certification of advanced proficiency in English and can be employed for academic and professional purposes (English Language Institute, 2006). Although the full-length test includes multiple-choice and constructed-response items measuring all four language skills (reading, writing, listening, and speaking) as well as a structural component consisting of a grammar, vocabulary, and a cloze test, this study only examines the multiple-choice sections of the test. These include the listening and reading subtests and the three structure subtests. The study is based on item-level data and uses multiple group confirmatory factor analysis to test various levels of measurement invariance and structural invariance of the factor model. The invariance analyses will consider both full and partial invariance, where partial invariance is observed when only some of the target parameters exhibit cross-group equivalence. This approach offers a more rigorous analysis of the factor structure differences between the two proficiency groups and allows a more fine-grained examination of systematic changes in the factor structure in relation to differences in language proficiency.

Methods

Data

This study is based on data from 34,599 examinees who took the ECPE during the 2005–2006 administration. The test consists of a total of 150 multiple-choice items that include 50 listening comprehension items (15 short conversations, 20 question items, 15 radio interviews), 30 sentence-completion items targeting grammar knowledge, 20 cloze test items based on a single reading passage, 30 vocabulary items, and 20 reading comprehension questions based on four reading passages.

Because no independent information was available to divide the examinee sample into low- and high-proficient groups, it was necessary to split the test items into two test halves of odd- and even-numbered items, keeping the same number of items per subtest within each test half. A total score was computed for each examinee based on the odd-numbered test half. This score is used to partition the sample into approximate halves of the top and bottom scoring examinees yielding accordingly the low and high proficient examinee groups. The analyses of measurement and structural invariance were then carried out on the test half with even-numbered items, henceforth ECPE (even). This procedure permitted the use of independent information to determine the proficiency groups when no external test data was available. Because the ECPE consists of a total of 150 items with 20 to 50 items per subtest, the test was deemed large enough to ensure that each test half would be an adequate representation of the full-length test. It should also be noted that in order to maintain equal proportions of items per activity type in the two test halves, two listening comprehension items were excluded, one randomly selected short conversation item (item 9) and one randomly selected radio interview question (item 42).

Table 1 provides descriptive information on the score distributions of the low and high proficiency groups and the total sample in both test halves. In order to evaluate whether the splitting of the test items yielded approximately equivalent score distributions in both test halves, a dependent-samples *t*-test was conducted using the entire examinee sample, which compared the mean difference of the total scores from each test half. The results are provided at the bottom of Table 1. Although a statistically significant mean difference was found, this is mostly a reflection of the extremely large sample size used in this analysis. When an effect size measure was computed, it indicated that the mean difference was very small ($d = 0.15$). It is therefore concluded that the score distributions on the two test halves are approximately equivalent. In order to determine if the difference in the mean scores achieved by each proficiency group is statistically significant, an independent-samples *t*-test was carried out using the total score of the ECPE (even). The results are reported in Table 2. As is shown, the mean difference between both proficiency groups is statistically significant, corresponding to a difference of more than one standard deviation of the total sample variance.

Table 1. Descriptive Statistics for Odd- and Even-Numbered Test Halves

Proficiency group	N	ECPE (odd) Mean	ECPE (odd) Std.	ECPE (even) Mean	ECPE (even) Std.
Low	17,309 (50.03%)	43.67	5.21	45.55	6.37
High	17,290 (49.97%)	56.87	4.48	56.46	6.00
Total	34,599 (100%)	50.27	8.19	51.0	8.25
T-test*	mean diff. = 0.73, t = 27.68, *df* = 34598, p < .001, effect size d = 0.15, $r_{odd\text{-}even}$ = 0.821				

* Dependent samples t-test for total sample

Table 2. Comparisons of Proficiency Groups for ECPE (even)

	Mean diff.	Observed t*	*df.*	Sig.	Effect size d
Low vs.high	10.91	163.905	34479.364	< .001	1.32

* Independent samples t-test, equal variances not assumed

Procedures

Invariance tests of factor models are generally conducted within the framework of multiple group confirmatory factor analysis and involve a sequence of logically ordered nested model comparisons (Jöreskog, 1971). These comparisons proceed in a stepwise fashion whereby increasingly more restrictive factor models are tested vis-à-vis a less restrictive baseline model. The cross-group invariance analyses in this study were carried out at various levels of the factor model and involve both measurement and structural aspects.

To examine the invariance of the measurement model the equivalence of factor loadings (metric invariance) and the equivalence of the item thresholds (scalar invariance) were tested in tandem. The simultaneous evaluation of metric and scalar invariance is necessary because the item probability curve of a categorical outcome variable is influenced by both parameters (Muthén & Muthén, 1998–2007). It should be noted that cross-group equivalence of the item residual variances was not examined because Delta parameterization was used in the specification of the invariance model. This parameterization is the recommended setting in Mplus 5 (Muthén & Asparouhov, 2002) and does not permit the estimation of residual variances as model parameters.

Because full measurement invariance rarely holds in empirical research, partial measurement invariance was also explored whenever full invariance was not obtained. Once partial metric and scalar invariance were determined—that is, at least one item per latent factor in addition to the marker item used for scale identification exhibited metric and scalar invariance (Byrne, Shavelson, & Muthén, 1989; Steenkamp & Baumgartner, 1998)—further analyses were carried out to examine the equivalence of the structural relations of the latent construct. Structural invariance involves the testing of the equivalence of factor variances and factor covariances (Steenkamp & Baumgartner, 1998; Vandenburg & Lance, 2000).

An important prerequisite for tests of measurement and structural invariance to be meaningful is that the configuration of latent factors is constant across the comparison groups

(Byrne et al., 1989). This requires that the pattern of salient and nonsalient factor loadings are equivalent (Steenkamp & Baumgartner, 1998). In this study, configural invariance was examined by comparing the fit of three hypothesized factor models separately for each proficiency group. The selection of the three factor models was guided by the findings from an exploratory factor analysis (EFA) and by substantive considerations.

In this study, all invariance analyses were carried out on the ECPE (even) using the software Mplus 5 (Muthén & Muthén, 1998–2007) with the WLS estimator. This estimator has been developed for analyses based on polychoric correlations from categorical outcome variables and has been shown to perform adequately with large sample sizes (Flora & Curran, 2004). According to Jöreskog and Sörbom (1996), the suggested minimum sample size for this estimator is (k+1)(k+2)/2 or 2850 examinee responses, where *k* is the number of model indicators. The sample sizes used in this study are 17,309 and 17,290, respectively for the low and high proficiency groups, and far exceed the recommended minimum.

Results

Exploratory Factor Analysis

The purpose of the exploratory factor analysis is to determine the number of latent factors that best account for the relationships among the observed variables of the ECPE (even). Findings from this analysis were used to inform the baseline model specification for the invariance analyses. Solutions with up to seven latent factors were obtained using oblique rotation. This cut-off was chosen on the consideration that a factor configuration that represents each of the five subtests and two additional factors for each of the listening activities is theoretically feasible. In order to determine the number of latent factors the following criteria were considered: the shape of the scree plot, parallel analysis using a random data set of equal size and complexity (Horn, 1965), the Factor Difference Ratio Index (FDRI; based on Hattie, 1985), the Kaiser criterion as an upper bound for the number of factors to be retained (Kaiser, 1960), and finally the interpretability of the obtained factor loading patterns and their suitability to obtain simple structure. In order to conduct the parallel analysis, principal components analysis was performed first using SPSS 15 to extract the eigenvalues and to compare them to the eigenvalues from an equivalent random data set. Then Mplus 5 with the WLS estimator was used to obtain the factor loading patterns for each of the seven factor solutions.

The extraction of eigenvalues from the initial principal components analysis, which is summarized in Table 3, indicates the presence of a strong first factor and multiple secondary factors. According to the scree plot presented in Figure 1, there appear to be two break points suggesting the retention of either one factor accounting for 7.4% of the variance or four factors accounting for a total of 14.8% of the variance. In order to evaluate the relative strength of the first factor, the FDRI was computed by dividing the difference between the first and second eigenvalue by the difference between the second and third eigenvalue (Michigan English Language Assessment Battery Technical Manual, 2003). This yielded a value of 9.96, which exceeds the critical value of three, thus supporting a one-factor solution. It should be noted that the Kaiser criterion and results from the parallel analysis suggested the retention of more than seven factors, which is more than would be supported by theory and was therefore not considered further. The scree plot of the extracted eigenvalues from the random data set is also given in Figure 1.

Table 3. Eigenvalue Extraction from ECPE (even)

	Eigenvalues		Cumulative % of
Factors	Random data	ECPE data	variance explained
1	1.091	5.488	7.42
2	1.087	2.115	10.27
3	1.086	1.776	12.67
4	1.080	1.575	14.80
5	1.073	1.290	16.54
6	1.071	1.160	18.11
7	1.066	1.095	19.59
8	1.060	1.072	21.04
9	1.058	1.050	22.46
10	1.055	1.038	23.86

An examination of the factor loading patterns suggests that a correlated three-factor solution exhibits the most interpretable structure with a separate listening and vocabulary factor and a third factor combining the grammar, cloze, and reading items. Other factor solutions exhibited either undefined factors or too many cross-loading items. The factor loading patterns along with the factor correlation matrices are provided in Appendix A. Overall, the findings from the exploratory factor analysis indicate that both a one-factor model and a correlated three-factor model appear to provide good model fit. Both models will therefore be considered as baseline factor models for the invariance analyses. In addition, a correlated five-factor model that represents the subtest structure of the ECPE will also be considered, because this is the structure that was targeted during test construction. All three baseline models will be specified with simple structure imposed.

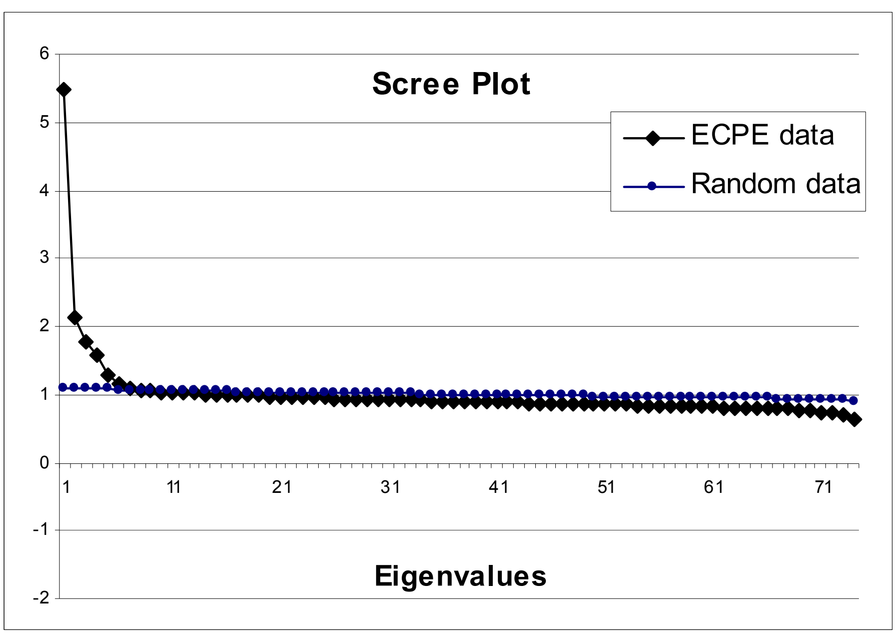

Figure 1. Scree Plot of ECPE (even) and Equivalent Random Data Set

Comparison of Baseline Models

In order to determine the appropriate baseline model for each proficiency group, the three hypothesized factor models were estimated and compared separately for each group. Because the one- and three-factor model are each nested in the five factor model, chi-square difference tests were computed to evaluate the improvement in model fit for the less restrictive five-factor model. In addition, several model fit indices were considered. The X^2/df ratio is an index intended for analyses with larger samples where the X^2 statistic would otherwise be too powerful. Values around 5.0 and below have been suggested as indications of reasonable model fit (Bollen, 1989). However, other authors recommend values of less than 2.0 (e.g., Ullman, 2001). The RMSEA and CFI fit indices were also considered with the following values chosen as indicators of good model fit: values less than 0.05 for the RMSEA, and values greater than 0.95 for the CFI (Hu & Bentler, 1999).

Table 4 presents the results from the chi-square difference tests and Table 5 provides the fit indices for all factor models in each group. The chi-square difference tests between the one- and five-factor model and between the three- and five-factor model were significant in each group. This indicates that the additional constraints specified in the one- and three-factor model result in a significant drop in model fit relative to the fit of the five-factor model. This conclusion is also supported by the other fit indices, which indicate improvement in model fit from the one-factor model to the five-factor model. In terms of the model fit of the individual models, only the criterion value of the RMSEA is met in all factor models. The X^2/df ratio supports the five factor model in both proficiency groups when the criterion value of 5.0 is applied but not under the more stringent criterion value of 2.0 or less. The CFI is far below the acceptable value of 0.95, therefore, not in support of any of the factor models. The discrepancy between the CFI value and the RMSEA may be a result of the very large sample size, the low to moderate factor loadings, and the relative complexity of the factor model, all of which have been reported to be less suitable conditions for the CFI (Beauducel & Wittman, 2005; Rigdon, 1996). Because of the discrepancy found between the CFI and RMSEA, it was decided to also consider McDonald's Fit Index (MFI; McDonald & Marsh, 1990), which, unlike the CFI and RMSEA, is not a comparative index of model fit, but only considers the model chi-square, model degrees of freedom, and sample size (Ullman, 2001). MFI values of 0.9 or higher are considered acceptable (Worthington & Whittaker, 2006). The MFI results for this analysis are given in Table 5. As is shown, the MFI also doesn't meet the criterion value of 0.9, but the results are much closer to the acceptable criterion than the CFI. Given these findings, it was concluded that the five-factor model provides the best model fit in both proficiency groups relative to the other factor models and that the five-factor model exhibits acceptable model fit as indicated by the RMSEA and the X^2/df ratio.

Table 4. Chi-square Difference Tests for Hypothesized Factor Models

	X^2	*df*	X^2-diff.	*df*-diff.	Sig.
High proficient					
5-factor	13098.3	2617			
3-factor	14424.6	2624	1326.3	7	< .001
1-factor	16997.4	2627	3899.1	10	< .001
Low proficient					
5-factor	11842.1	2617			
3-factor	13468.9	2624	1626.8	7	< .001
1-factor	16508.0	2627	4665.9	10	< .001

Table 5. Factor Model Fit Indices for Low and High Proficient Examinees

	RMSEA	CFI	X^2/df	MFI
High proficient				
5-factor	0.015	0.587	5.0	0.859
3-factor	0.016	0.535	5.5	0.843
1-factor	0.018	0.434	6.5	0.812
Low proficient				
5-factor	0.014	0.584	4.5	0.875
3-factor	0.015	0.511	5.1	0.855
1-factor	0.017	0.375	6.3	0.818

Invariance Analyses

Because identical baseline factor models were found for each group, this established the configural invariance of the ECPE (even) across proficiency levels. However, configural invariance does not ensure that the test items measure the test construct equally across groups. Therefore, analyses of measurement invariance were carried out next. Measurement invariance is examined by testing the equivalence of factor loadings (metric invariance) and item thresholds (scalar invariance) using multiple group confirmatory factor analysis. First the correlated five-factor model was specified with no cross-group equality constraints imposed on the factor loadings and item thresholds. This baseline model reflects the situation of complete measurement noninvariance. Then metric and scalar invariance were tested simultaneously by setting equality constraints on both factor loadings and item thresholds. For the purpose of model identification, the loadings of the best-defined item indicators per latent factor were constrained to 1. Using chi-square difference tests it was then possible to statistically evaluate whether the equality constraints significantly worsened the fit of the full measurement invariance model. Table 6 provides the results of this analysis.

Table 6. Chi-Square Difference Tests for Full Measurement and Structural Invariance

Invariance factor model	X^2	*df*	X^2-diff.	*df*-diff.	Sig.
Measurement invariance					
Full non-invariance (baseline)	24940.38	5234			
Full invariance	26369.11	5298	1428.73	64	<.001
Structural invariance					
Partial measurement (baseline)	25819.48	5303			
Variance/Covariance invariance	26410.42	5318	590.934	15	<.001

As is shown, the chi-square difference test was statistically significant, indicating a significant drop in model fit for the full measurement invariance model in comparison to the baseline noninvariance model. In order to investigate the degree of partial measurement invariance, modification indices were examined to identify the items with noninvariant parameters. The equality constraints were then removed successively from each item with the largest, statistically significant model chi-square drop and the invariance model was then re-estimated. This process was repeated until no more items were indicated by the modification indices. During this procedure, Holm's modified Bonferroni correction was applied in order to control for the familywise error rate (Holm, 1979). A total of 34 items out of a total of 74 in the ECPE (even) exhibited measurement noninvariance. The vast majority of noninvariant items were found in the listening, grammar, and vocabulary subtests, but only two items in the reading subtest and none of the items in the cloze subtest were found to be noninvariant. The noninvariant items are listed by subtest in Table 7. Final estimates of factor loadings and item thresholds are provided in Appendix B.

Table 7. Items Exhibiting Measurement Noninvariance

Subtest	N	Percent	Items
Listening	14/24	58%	2, 6, 8, 18, 20, 28, 30, 32, 36, 38, 40, 46, 48, 50
Grammar	10/15	67%	54, 56, 66, 68, 70, 72, 74, 76, 78, 80
Cloze	0/10	0%	-
Vocabulary	9/15	60%	108, 110, 112, 114, 116, 118, 120, 122, 128
Reading	2/10	20%	134, 136

Because partial measurement invariance is sufficient to carry out invariance analyses of the factor variances and covariances (Steenkamp & Baumgartner, 1998), these analyses were conducted next. The result of the chi-square difference test is also given in Table 6. Note that the new baseline model to which the variance/covariance model is compared is the partial measurement invariance model with the equality constraints removed for 34 of the 74 factor indicators. The test of factor variance/covariance equivalence resulted in a statistically significant chi-square difference indicating noninvariance of the factor variances and covariances. After examining modification indices to identify individual parameters that are invariant, none of the variance and covariance parameters in the five-factor model exhibited cross-group equivalence. Table 8 provides the factor correlation matrices and factor variances for each proficiency group.

Table 8. Factor Variances and Factor Correlations by Proficiency Group

Low	Variance	Listening	Grammar	Cloze	Vocabulary
Listening	0.086				
Grammar	0.087	0.453			
Cloze	0.048	0.455	0.816		
Vocabulary	0.125	0.264	0.463	0.378	
Reading	0.215	0.352	0.37	0.483	0.322
High					
Listening	0.100				
Grammar	0.144	0.626			
Cloze	0.058	0.608	0.888		
Vocabulary	0.115	0.584	0.54	0.528	
Reading	0.267	0.449	0.425	0.597	0.397

Low proficiency: avg. factor correlation $r = 0.44$, std. = 0.15; High proficiency: avg. factor correlation $r = 0.56$, std. = 0.14.

Table 8 shows that all but one factor variance and all factor correlations are smaller in the low proficiency group than in the high proficiency group. On average, the subtest factors in the high proficiency group correlate with each other at around 0.56 (Std.=0.14), which according to Cohen's criteria indicates a strong correlation, whereas in the low proficiency group the average factor correlation is a moderate 0.44 (Std.=0.15) (Cohen, 1988). The higher magnitude of the correlations of the high proficient examinees suggests a somewhat more cohesive and convergent factor structure for this group, and therefore indicates a decreasing trend in the relationship between language proficiency and factor differentiation. However, a comparison of the noninvariant factor loadings across the two groups did not produce further evidence that the factor structure becomes more differentiated as examinee's proficiency improves. There were approximately equal numbers of factor loadings that were either above or below the loading of the other group (see Appendix B), therefore, providing no indication that the latent factors are more salient in one group than in the other.

Some of the largest discrepancies in the correlation coefficients can be found between the listening factor and the grammar, cloze, and vocabulary factors. In each case, the correlations differ by 0.15 or more. These discrepancies may be an indication that listening ability, at least initially, develops differently from the other language skills, in particular the basic skills measured by the grammar, cloze, and vocabulary tests. In both groups, the rather high correlation between the grammar and cloze subtest stands out. This high correlation is consistent with findings from other studies which concluded that cloze tests are closely related if not identical to tests of grammatical knowledge (e.g., Purpura, 1999; Saito, 2003). Interestingly though, the exploratory factor analysis did not indicate a clear convergence of the two subtests beyond the three-factor solution, in which the cloze items loaded together with the grammar and the reading items. This suggests that grammar knowledge alone does not fully account for the covariance pattern of the cloze items. Rather cloze items appear to also tap into other aspects of language proficiency.

Discussion

The goal of this study was to investigate the degree of factor differentiation in the ECPE across two examinee groups with different language proficiency levels. The study was motivated by inconsistencies in the research literature, that had reported contradictory evidence regarding the relationship between language proficiency and factor differentiation. In addition, the study attempted to address methodological shortcomings of previous research by examining item-level data and by using tests of nested model comparisons. This allowed a more thorough and rigorous evaluation of factorial differences using statistical criteria.

One of the main findings in this research study is that the factorial structure of the ECPE differs across the low and high proficiency group. Specifically, the study found that the measurement properties of a substantial number of items as well as the factor variances and factor correlations are noninvariant. The noninvariance of the structural model (the factor variances and correlations) suggests that group differences exist in the underlying latent construct measured by the five ECPE subtests. The higher factor correlations in the high proficiency group point to a decreasing relationship between factor differentiation and language proficiency. In other words, as ECPE examinees become more proficient in English, individual skills develop more uniformly with other skills and converge in a more general language proficiency configuration.

The results confirm findings in Oltman et al. (1988) and in Kunnan (1992), who both report higher salience of individual factors in the low proficiency group and conclude that language proficiency is more differentiated at earlier stages of language acquisition. Other studies, such as Ginther and Stevens (1998), however, found evidence for an increasing relationship. A possible reason for the difference in the findings is that Ginther and Stevens compared native and bilingual speakers to classroom language learners. One of their main findings is that the factor correlations are not only lower in the highly proficient native speaker groups but Speaking in particular emerges as a highly distinct and almost unrelated factor. The explanation offered by the authors was that native and bilingual speakers have very different language learning experiences from classroom learners, which leads native speakers to develop primarily strong oral proficiency skills whereas the classroom learners are constrained by what is offered in the academic environment. While Ginther and Stevens compared groups with very different language learning experiences, the present study focused only on proficiency differences in nonnative speakers with presumably mostly classroom learning experiences. In addition, this study does not include a Speaking subtest but focused primarily on skills emphasized in formal instruction with the possible exception of Listening. It is therefore plausible that the selection of the particular examinee groups, especially of the high proficiency examinees, and the composition of the test instrument have influenced the respective outcomes of each study. Other authors have made similar observations and pointed to examinee characteristics such as amount of language exposure and type of instruction as likely sources that influence the factor structures of language tests (e.g., Kunnan, 1995; Sang, Schmitz, Vollmer, Baumert, & Roeder., 1986).

In addition to the question of factor differentiation in language tests, this study also addresses important aspects of the construct validity of the ECPE. The measurement invariance analysis reveals that the number of items with unequal measurement properties is substantial (i.e., 46%). Differences were mostly found in the listening, grammar, and vocabulary subtests but not in the cloze tests and only in two items of the reading subtest. All of the subtests had at least five or more invariant items including the marker item defining the factor scale, thus meeting the requirement of partial measurement invariance. Generally, partial measurement invariance is considered sufficient for test scores to be comparable, but the reliability of estimating the true score is affected by the number of noninvariant items (Millsap & Kwok, 2004). Because some of the subtests in this study had high proportions of noninvariant items (e.g., grammar), this likely affects the accuracy by which these subtests measure language proficiency.

While partial measurement invariance does not affect the comparability of scores from individual subtests as long as these are unidimensional, the noninvariance of the structural model has implications for the comparability of any composite scores that may be formed on the basis of a latent variable model. Because the factor correlations in the ECPE differ between the two groups, the implicit weighting of the subtest scores in a composite score would also change. As a consequence, composite scores from examinees with low proficiency carry a different meaning than composite scores from high proficiency examinees. Alternatively, composites may be formed by a linear combination of individual subtest scores, for example through summation. However, given the lower factor correlations for the low proficiency examinees, linear composite scores for this examinee group would be less reliable measures of general language proficiency than linear composites for high proficiency examinees. This limits the utility of these scores, for example for placement decisions where accuracy at the lower end of proficiency is particularly important.

Study Limitations

There are several limitations of this study that affect the generalizability and interpretation of the research findings. Foremost, it is important to emphasize that the results of the invariance analyses are based on only half of the multiple choice items of the ECPE. Although the odd/even split of test items is a fairly common procedure used to obtain symmetrical halves from a single test, it does not guarantee that the test halves are mirror images of each other, and that they are accurate representations of the full test. In order to have more confidence in the results of this study, the analyses should therefore be cross-validated on the odd-numbered test half.

An important prerequisite for invariance analyses is that the configuration of the baseline models are appropriate representations of the underlying trait structure. The chi-square difference tests indicate that the correlated five-factor model provides the best overall fit when compared to two other models. However, the values of the model fit indices are not consistent. RMSEA and X^2/df ratio both meet the criterion values of model fit, but MFI and, especially, CFI do not. Discrepancies between model fit statistics generally cast doubt on the adequacy of the model. However, Beauducel and Wittman (2005) also point out that in certain modeling contexts similar to this study the CFI tends to be overly sensitive to minor distortions in simple structure, which led these authors to recommend against the use of the CFI in contexts where small deviations from simple structure are to be expected. This caveat notwithstanding, without converging results from various model fit indices, we cannot be completely confident about the appropriateness of the five-factor model. Cross-validation on the odd-numbered test half may provide more information regarding the appropriateness of the five-factor model.

The stated goal of this study was to compare the factor structures produced by different proficiency level groups in order to examine changes in the degree of factor differentiation. To observe these changes, however, only two proficiency groups were compared. This was necessitated by the large sample size requirement of the WLS estimator used in the analyses. The limitation of this is that a two-group comparison provides only a crude indication of the underlying trend in factor differentiation, whereas comparisons of more than two proficiency levels would have allowed a more detailed evaluation of systematic changes. Furthermore, the determination of the two proficiency groups was essentially based on an arbitrary cut point with no external reference that could link the low and high proficiency group to a corresponding proficiency level in another study. This limits the comparability of the study's findings with other research. Future studies should therefore consider using external measures with established performance level categories in order to facilitate comparability among studies.

Acknowledgements

I would like to thank the English Language Institute of the University of Michigan for providing the funding for this research study. I am also indebted to my advisor James Bovaird at the University of Nebraska, Lincoln, who is always available when advise is dearly needed and who never fails to encourage me in my research efforts.

References

Beauducel, A., & Wittman, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling*, *12*(1), 41–75.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structure: the issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456–466.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum

English Language Institute. (2006). *Examination for the certificate of proficiency in English: Information bulletin*. Ann Arbor: English Language Institute, University of Michigan.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466–491.

Ginther, A., & Stevens, J. (1998): Language background, ethnicity, and the internal construct validity of the advanced placement Spanish language examination. In A. J. Kunnan (ed.), *Validation in language assessment* (pp. 169–194). Mahwah, NJ: Lawrence Erlbaum Associates.

Hattie, J. (1985). Methodology Review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, *9*, 139–164.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.
Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *32*, 179–185.
Hosley, D., & Meredith, K. (1976). Inter- and Intra-test correlates of the TOEFL. *TESOL Quarterly*, *13*, 209–217.
Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.
Johnson, J. S., Yamashiro, A., & Yu, J. (2003). *The ECPE report: 2002*. Ann Arbor: English Language Institute, University of Michigan.
Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409–426.
Jöreskog, K. G., & Sörbom, D. (1996). *LISREL8 user's reference guide*. Chicago: Scientific Software International.
Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141–151.
Kunnan, A. J. (1992). An investigation of a criterion-referenced test using G-theory, factor and cluster analyses. *Language Testing*, *9*, p.30–49.
McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin*, *107*, 247–255.
Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, *58*, 525–543.
English Language Institute. (2003). *Michigan English Language Assessment Battery Technical Manual 2003*. Ann Arbor: English Language Institute, University of Michigan.
Millsap, R. E., & Kwok, O-M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*, *9*(1), 93–115.
Muthén, B., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. Mplus Web Notes: No. 4. www.statmodel.com.
Muthén, L.K., & Muthén, B.O. (1998–2007). *Mplus User's Guide. Fifth Edition*. Los Angeles, CA: Muthén & Muthén
Nunnally, J. C. (1978). *Psychometric Theory.* New York: McGraw-Hill.
Oltman, P. K., Stricker, L. J., and Barrows, T. (1988). *Native language, English proficiency, and the structure of the TOEFL* (TOEFL research report 27). Princeton, NJ: Educational Testing Service.
Purpura, J. E. (1999). *Learner strategy use and performance on language tests: A structural equation modeling approach.* Cambridge, UK: Cambridge University Press.
Rigdon, E. E. (1996). CFI versus RMSEA: A comparison of two fit indexes for structural equation modeling. *Structural Equation Modeling*, *3*(4), 369–379.
Saito, Y. (2003). *Investigating the construct validity of the Cloze section in the Examination for the Certificate of Proficiency in English* (Spaan Fellow Working Papers in Second or Foreign Language Assessment, Volume 1). Ann Arbor: English Language Institute, University of Michigan.

Sang, F., Schmitz, B., Vollmer, H. J., Baumert, J., & Roeder, R. M. (1986). Models of second language competence: a structural equation approach. *Language Testing*, *3*(1), 54–79.

Shin, S.-K. (2005). Did they take the same test? Examinee language proficiency and the structure of language tests. *Language Testing*, *22*(1), 31–57.

Steenkamp, J. B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*, 78–90.

Swinton, S. S., & Powers, D. E. (1980). *Factor analysis of the TOEFL (*TOEFL Research Reports 6). Princeton, NJ: Educational Testing Service.

Ullman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell, *Using multivariate statistics* (pp. 653–771). Needham Heights, MA: Allyn and Bacon.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: a content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*, 806–838.

Appendix A

The following Tables A1 through A20 present the factor model solutions from the exploratory factor analysis (all loadings below 0.2 are displayed in grey). Table 21 presents the corresponding factor correlation matrices obtained from each factor solution.

Table A1. One-, Two-, and Three-Factor Solution for Listening

	F1	F1	F2	F1	F2	F3
L2	0.396	0.494	-0.077	0.434	-0.071	0.030
L4	0.314	0.378	-0.046	0.385	0.023	0.083
L6	0.345	0.458	-0.095	0.352	-0.176	-0.076
L8	0.436	0.642	-0.217	0.536	-0.134	-0.120
L10	0.274	0.326	-0.033	0.295	-0.049	0.038
L12	0.329	0.413	-0.075	0.409	0.028	0.075
L14	0.213	0.163	0.084	0.150	-0.025	0.128
L16	0.352	0.410	-0.030	0.338	-0.114	0.017
L18	0.428	0.503	-0.052	0.433	-0.097	0.040
L20	0.384	0.470	-0.071	0.416	-0.064	0.035
L22	0.397	0.571	-0.175	0.479	-0.123	-0.089
L24	0.387	0.434	-0.028	0.412	-0.006	0.110
L26	0.309	0.235	0.103	0.238	0.005	0.190
L28	0.366	0.393	0.000	0.386	0.018	0.143
L30	0.346	0.397	-0.018	0.300	-0.185	-0.019
L32	0.346	0.357	0.020	0.356	0.014	0.152
L34	0.267	0.305	-0.006	0.232	-0.112	0.020
L36	0.414	0.367	0.113	0.233	-0.232	0.085
L38	0.260	0.113	0.198	0.096	-0.034	0.234
L40	0.335	0.248	0.135	0.205	-0.057	0.196
L44	0.398	0.428	0.009	0.402	-0.017	0.141
L46	0.540	0.379	0.244	0.378	0.036	0.429
L48	0.445	0.230	0.305	0.184	-0.061	0.376
L50	0.152	0.136	0.024	0.164	0.073	0.128

Table A2. Four- and Five-Factor Solution for Listening

	F1	F2	F3	F4	F1	F2	F3	F4	F5
L2	0.452	-0.038	-0.001	0.037	0.507	-0.115	0.058	0.014	0.106
L4	0.371	-0.027	0.003	0.032	0.369	0.051	-0.056	0.053	-0.036
L6	0.342	-0.188	0.051	-0.079	0.404	0.011	0.140	-0.065	0.065
L8	0.520	-0.117	0.098	-0.164	0.547	0.030	0.037	-0.151	0.104
L10	0.298	-0.065	0.004	0.013	0.339	-0.031	0.036	0.024	0.024
L12	0.411	0.005	-0.005	0.032	0.408	0.026	-0.072	0.029	-0.015
L14	0.161	-0.001	0.028	0.108	0.167	0.004	-0.023	0.108	0.028
L16	0.333	-0.143	0.012	0.020	0.364	0.032	0.080	0.028	0.022
L18	0.424	-0.163	-0.019	0.037	0.436	0.089	0.060	0.043	-0.035
L20	0.414	-0.080	0.018	0.010	0.449	-0.009	0.025	0.015	0.038
L22	0.461	-0.155	0.041	-0.108	0.457	0.131	0.036	-0.104	0.009
L24	0.399	-0.068	-0.024	0.093	0.349	0.162	-0.066	0.078	-0.069
L26	0.247	-0.060	-0.063	0.191	0.255	0.022	0.001	0.189	-0.057
L28	0.388	0.041	0.049	0.075	0.390	0.006	-0.098	0.070	0.044
L30	0.292	-0.152	0.076	-0.021	0.38	-0.053	0.150	-0.017	0.122
L32	0.362	0.055	0.051	0.089	0.343	0.024	-0.109	0.079	0.039
L34	0.231	-0.062	0.088	-0.008	0.243	0.037	0.015	-0.004	0.086
L36	0.236	-0.022	0.249	0.030	0.286	-0.053	0.027	0.040	0.274
L38	0.112	0.027	0.065	0.204	0.077	0.065	-0.076	0.196	0.046
L40	0.213	0.047	0.119	0.143	0.187	0.056	-0.099	0.134	0.093
L44	0.410	0.085	0.121	0.058	0.400	0.005	-0.151	0.053	0.113
L46	0.398	0.239	0.196	0.302	0.283	0.147	-0.358	0.266	0.121
L48	0.211	0.127	0.176	0.304	0.159	0.072	-0.190	0.284	0.142
L50	0.164	0.135	0.061	0.071	0.076	0.125	-0.209	0.048	0.000

Table A3. Six-Factor Solution for Listening

	F1	F2	F3	F4	F5	F6
L2	0.530	-0.011	0.008	0.057	0.041	-0.052
L4	0.299	0.183	0.052	-0.035	-0.017	0.010
L6	0.397	-0.032	0.150	-0.031	0.026	-0.003
L8	0.544	-0.004	0.046	-0.107	0.032	0.076
L10	0.270	0.087	0.100	-0.008	0.014	-0.021
L12	0.346	0.166	0.018	-0.037	-0.015	0.017
L14	0.173	0.058	-0.027	0.097	0.002	0.029
L16	0.332	0.048	0.113	0.020	-0.005	0.035
L18	0.334	0.162	0.185	-0.034	-0.018	0.037
L20	0.473	0.008	0.000	0.050	-0.032	0.041
L22	0.413	0.064	0.114	-0.117	-0.015	0.120
L24	0.302	0.170	0.037	0.011	-0.086	0.151
L26	0.192	0.167	0.076	0.109	-0.048	0.000
L28	0.295	0.228	0.033	-0.041	0.072	-0.023
L30	0.417	-0.081	0.101	0.052	0.055	-0.016
L32	0.244	0.230	0.035	-0.032	0.064	-0.011
L34	0.283	-0.023	-0.003	0.034	0.037	0.067
L36	0.383	-0.093	-0.079	0.133	0.171	0.050
L38	0.057	0.145	-0.022	0.131	0.045	0.058
L40	0.185	0.124	-0.063	0.095	0.072	0.070
L44	0.343	0.204	-0.042	-0.027	0.105	0.010
L46	0.139	0.487	-0.105	0.021	0.189	0.093
L48	0.075	0.317	-0.044	0.128	0.175	0.043
L50	0.018	0.212	-0.091	-0.057	0.026	0.106

Table A4. Seven-Ffactor Solution for Listening

	F1	F2	F3	F4	F5	F6	F7
L2	0.538	-0.023	-0.027	0.067	0.043	-0.043	-0.033
L4	0.323	0.165	0.029	-0.042	0.007	0.015	0.039
L6	0.399	-0.049	0.131	-0.012	0.004	-0.014	-0.049
L8	0.563	-0.044	0.005	-0.084	0.057	0.066	-0.021
L10	0.252	0.094	0.106	-0.026	0.016	-0.036	-0.023
L12	0.356	0.151	0.008	-0.046	0.024	0.016	0.033
L14	0.182	0.065	-0.034	0.089	0.016	0.025	-0.003
L16	0.352	0.031	0.089	0.027	0.024	0.036	0.017
L18	0.346	0.145	0.171	-0.045	0.005	0.028	0.010
L20	0.482	0.000	-0.030	0.060	-0.019	0.029	-0.029
L22	0.430	0.029	0.087	-0.100	-0.005	0.107	-0.017
L24	0.326	0.161	0.017	-0.001	-0.037	0.141	0.035
L26	0.202	0.174	0.071	0.084	-0.007	0.003	0.046
L28	0.289	0.226	0.032	-0.067	0.066	-0.028	-0.014
L30	0.414	-0.088	0.070	0.072	0.045	-0.020	-0.057
L32	0.218	0.244	0.049	-0.070	0.040	-0.024	-0.044
L34	0.272	-0.020	-0.016	0.042	0.012	0.052	-0.075
L36	0.353	-0.068	-0.082	0.133	0.101	0.026	-0.155
L38	0.074	0.161	-0.026	0.107	0.087	0.058	0.042
L40	0.176	0.147	-0.062	0.070	0.058	0.056	-0.052
L44	0.312	0.221	-0.033	-0.058	0.052	-0.014	-0.093
L46	0.112	0.529	-0.078	-0.057	0.134	0.075	-0.075
L48	0.079	0.345	-0.039	0.074	0.171	0.041	-0.011
L50	0.013	0.219	-0.081	-0.084	0.005	0.091	-0.026

Table A5. One-, Two-, and Three-Factor Solution for Grammar

	F1	F1	F2	F1	F2	F3
G52	0.414	0.357	0.113	0.181	-0.268	0.064
G54	0.325	0.378	-0.016	0.248	-0.246	-0.083
G56	0.339	0.379	-0.019	0.329	-0.076	0.053
G58	0.352	0.304	0.096	0.166	-0.244	0.034
G60	0.389	0.424	0.019	0.245	-0.298	-0.053
G62	0.348	0.335	0.054	0.187	-0.255	-0.011
G64	0.240	0.272	-0.001	0.186	-0.152	-0.018
G66	0.372	0.213	0.234	0.029	-0.334	0.096
G68	0.376	0.221	0.238	-0.009	-0.440	0.015
G70	0.259	0.076	0.252	-0.030	-0.220	0.153
G72	0.475	0.457	0.077	0.308	-0.254	0.057
G74	0.392	0.340	0.130	0.003	-0.569	-0.160
G76	0.314	0.278	0.083	0.073	-0.284	0.004
G78	0.340	0.301	0.098	0.074	-0.457	-0.145
G80	0.419	0.391	0.107	0.089	-0.558	-0.19

Table A6. Four- and Five-Factor Solution for Grammar

	F1	F2	F3	F4	F1	F2	F3	F4	F5
G52	0.162	-0.206	0.137	0.076	0.088	0.315	0.032	0.06	0.045
G54	0.225	-0.341	-0.023	-0.026	0.284	0.123	0.246	-0.018	-0.021
G56	0.333	-0.133	-0.023	0.048	0.332	0.101	0.031	0.047	-0.039
G58	0.148	-0.332	-0.037	0.110	0.126	0.275	0.183	0.104	-0.098
G60	0.218	-0.248	0.132	-0.033	0.159	0.318	0.085	-0.046	0.038
G62	0.167	-0.232	0.094	0.020	0.093	0.312	0.062	0.006	0.009
G64	0.178	-0.199	-0.002	0.014	0.175	0.146	0.108	0.014	-0.031
G66	0.024	-0.303	0.092	0.157	0.004	0.256	0.179	0.148	0.034
G68	-0.018	-0.455	0.046	0.124	0.049	0.170	0.369	0.136	0.037
G70	-0.023	-0.240	-0.001	0.223	-0.008	0.137	0.177	0.224	-0.021
G72	0.292	-0.305	0.010	0.104	0.292	0.209	0.164	0.101	-0.023
G74	-0.024	-0.478	0.195	-0.060	-0.013	0.316	0.336	-0.053	0.130
G76	0.049	-0.186	0.171	0.024	-0.107	0.464	-0.028	-0.012	0.019
G78	0.057	-0.530	-0.012	0.003	0.132	0.210	0.427	0.012	-0.020
G80	0.062	-0.586	0.058	-0.047	0.168	0.181	0.490	-0.043	0.081

Table A7. Six-Factor Solution for Grammar

	F1	F2	F3	F4	F5	F6
G52	0.034	0.112	0.149	-0.001	0.047	0.260
G54	0.201	0.012	0.336	-0.034	-0.004	0.039
G56	0.268	0.123	0.123	-0.007	-0.044	0.072
G58	0.014	0.116	0.335	0.026	-0.061	0.167
G60	0.024	0.150	0.287	-0.154	0.094	0.198
G62	0.076	0.040	0.134	-0.006	-0.015	0.279
G64	0.098	0.076	0.209	-0.035	-0.011	0.083
G66	-0.059	0.065	0.278	0.097	0.050	0.177
G68	-0.073	0.026	0.491	0.082	0.095	0.031
G70	-0.080	0.088	0.263	0.161	0.010	0.056
G72	0.193	0.126	0.298	0.030	-0.007	0.124
G74	-0.060	-0.100	0.411	-0.028	0.139	0.216
G76	-0.068	0.000	0.011	0.003	-0.034	0.465
G78	0.074	-0.112	0.470	0.041	-0.013	0.122
G80	0.129	-0.170	0.497	0.029	0.069	0.094

Table A8. Seven-Factor Solution for Grammar

	F1	F2	F3	F4	F5	F6	F7
G52	0.069	0.099	0.133	-0.011	0.074	0.253	0.016
G54	0.208	-0.009	0.326	-0.029	-0.001	0.039	-0.003
G56	0.273	0.117	0.113	-0.016	-0.024	0.063	0.009
G58	0.042	0.103	0.329	0.016	-0.003	0.165	0.066
G60	-0.006	0.151	0.314	-0.183	0.02	0.173	-0.104
G62	0.082	0.044	0.134	-0.014	-0.026	0.252	-0.052
G64	0.104	0.067	0.206	-0.044	-0.009	0.075	-0.004
G66	-0.018	0.058	0.263	0.087	0.110	0.186	0.070
G68	-0.055	0.016	0.489	0.072	0.112	0.046	0.032
G70	-0.037	0.085	0.245	0.147	0.089	0.077	0.112
G72	0.202	0.122	0.296	0.012	0.000	0.107	-0.016
G74	-0.072	-0.105	0.415	-0.024	0.080	0.206	-0.095
G76	-0.036	-0.007	-0.008	0.007	-0.018	0.448	-0.025
G78	0.103	-0.142	0.438	0.058	0.046	0.137	0.047
G80	0.103	-0.177	0.494	0.048	0.012	0.085	-0.105

Table A9. One-, Two-, and Three-Factor Solution for Cloze

	F1	F1	F2	F1	F2	F3
C82	0.110	0.090	0.046	-0.058	-0.236	-0.075
C84	0.468	0.394	0.141	0.244	-0.263	0.100
C86	0.297	0.246	0.102	0.061	-0.303	-0.002
C88	0.041	0.043	0.012	-0.049	-0.144	-0.069
C90	0.391	0.327	0.127	0.097	-0.353	0.017
C92	0.632	0.526	0.208	0.201	-0.503	0.061
C94	0.425	0.357	0.128	0.174	-0.297	0.061
C96	0.270	0.234	0.064	0.162	-0.109	0.081
C98	0.277	0.255	0.064	0.114	-0.222	-0.003
C100	0.066	0.033	0.042	0.056	0.008	0.053

Table A10. Four- and Five-Factor Solution for Cloze

	F1	F2	F3	F4	F1	F2	F3	F4	F5
C82	-0.075	-0.114	0.169	-0.053	-0.145	0.231	0.024	-0.062	0.094
C84	0.236	-0.246	0.072	0.127	0.239	0.162	0.142	0.124	0.047
C86	0.043	-0.181	0.180	0.024	-0.003	0.235	0.074	0.011	0.116
C88	-0.055	-0.071	0.105	-0.056	-0.086	0.125	0.020	-0.063	0.065
C90	0.080	-0.120	0.317	0.001	-0.034	0.351	-0.041	-0.018	0.197
C92	0.180	-0.333	0.292	0.084	0.010	0.617	0.038	0.042	0.105
C94	0.166	-0.180	0.195	0.058	0.116	0.254	0.040	0.047	0.121
C96	0.162	0.044	0.184	0.026	0.098	0.130	-0.120	0.010	0.134
C98	0.102	-0.129	0.153	-0.006	0.014	0.288	-0.002	-0.024	0.062
C100	0.054	-0.031	-0.038	0.053	0.068	-0.002	0.025	0.054	-0.031

Table A11. Six-Factor Solution for Cloze

	F1	F2	F3	F4	F5	F6
C82	-0.070	-0.108	-0.016	0.007	0.042	0.254
C84	0.136	0.130	0.283	0.043	0.078	0.075
C86	0.027	-0.040	0.080	0.039	0.080	0.228
C88	-0.022	-0.102	-0.036	0.003	0.021	0.154
C90	0.071	-0.078	-0.082	0.055	0.098	0.411
C92	0.018	0.053	0.134	0.033	0.046	0.586
C94	0.159	-0.009	0.052	0.081	0.054	0.265
C96	0.189	-0.022	-0.166	0.065	0.051	0.205
C98	-0.040	0.091	0.111	-0.077	0.077	0.232
C100	0.007	0.085	0.093	-0.001	0.007	-0.046

Table A12. Seven-Factor Solution for Cloze

	F1	F2	F3	F4	F5	F6	F7
C82	-0.036	-0.124	-0.041	0.031	0.052	0.254	-0.012
C84	0.090	0.160	0.312	0.000	0.015	0.050	-0.108
C86	0.000	-0.017	0.093	0.029	0.005	0.199	-0.139
C88	-0.011	-0.108	-0.048	0.019	0.016	0.150	-0.026
C90	0.070	-0.062	-0.091	0.059	0.041	0.375	-0.145
C92	0.064	0.044	0.104	0.031	0.067	0.566	-0.037
C94	0.145	0.012	0.048	0.072	-0.001	0.234	-0.130
C96	0.128	0.026	-0.148	0.042	-0.051	0.163	-0.200
C98	-0.043	0.090	0.120	-0.095	0.053	0.216	-0.049
C100	-0.003	0.089	0.101	-0.013	-0.003	-0.045	-0.004

Table A13. One-, Two-, and Three-Factor Solution for Vocabulary

	F1	F1	F2	F1	F2	F3
V102	0.380	0.021	0.435	-0.094	-0.191	0.373
V104	0.378	0.363	0.075	0.226	-0.315	-0.043
V106	0.361	-0.012	0.455	-0.173	-0.308	0.289
V108	0.368	-0.037	0.488	-0.151	-0.239	0.357
V110	0.421	-0.049	0.556	-0.086	-0.088	0.537
V112	0.313	0.042	0.350	-0.090	-0.277	0.215
V114	0.575	0.216	0.460	0.098	-0.18	0.463
V116	0.468	0.359	0.179	0.255	-0.206	0.168
V118	0.315	0.071	0.310	0.011	-0.147	0.260
V120	0.528	0.105	0.502	0.162	0.093	0.642
V122	0.297	-0.112	0.482	-0.217	-0.225	0.335
V124	0.252	-0.199	0.526	-0.230	-0.089	0.449
V126	0.234	0.009	0.280	-0.091	-0.183	0.189
V128	0.345	0.114	0.285	0.127	0.059	0.408
V130	0.378	0.178	0.282	-0.011	-0.359	0.117

Table A14. Four- and Five-Factor Solution for Vocabulary

	F1	F2	F3	F4	F1	F2	F3	F4	F5
V102	-0.061	-0.086	0.100	0.402	-0.087	0.119	0.021	0.394	0.076
V104	0.211	-0.426	-0.038	0.026	0.324	0.053	0.388	0.041	0.002
V106	-0.145	-0.208	0.100	0.360	-0.049	-0.068	0.230	0.385	0.150
V108	-0.122	-0.226	-0.006	0.440	-0.039	-0.044	0.229	0.461	0.042
V110	-0.042	0.029	0.080	0.532	-0.026	-0.030	-0.038	0.530	0.094
V112	-0.079	-0.243	0.047	0.293	-0.021	0.049	0.219	0.303	0.060
V114	0.123	-0.075	0.117	0.462	0.095	0.120	-0.023	0.459	0.084
V116	0.256	-0.217	0.018	0.198	0.301	0.057	0.148	0.201	0.034
V118	0.031	-0.099	0.050	0.275	0.046	0.056	0.065	0.272	0.047
V120	0.206	0.098	-0.060	0.620	0.147	0.061	-0.185	0.598	-0.079
V122	-0.190	-0.186	0.013	0.422	-0.118	-0.038	0.202	0.440	0.054
V124	-0.199	-0.003	0.031	0.494	-0.181	-0.038	0.017	0.498	0.053
V126	-0.077	-0.142	0.038	0.247	-0.055	0.048	0.114	0.249	0.048
V128	0.142	0.160	0.088	0.345	0.018	0.186	-0.269	0.310	0.001
V130	0.002	-0.283	0.113	0.170	0.112	-0.037	0.287	0.195	0.159

Table A15. Six-Factor Solution for Vocabulary

	F1	F2	F3	F4	F5	F6
V102	0.013	0.004	-0.064	0.420	-0.006	0.186
V104	0.226	-0.035	0.467	0.027	0.033	-0.054
V106	0.039	-0.099	0.097	0.457	0.079	-0.022
V108	0.027	-0.045	0.116	0.503	-0.015	-0.013
V110	-0.057	0.232	-0.002	0.414	0.109	-0.034
V112	-0.055	0.031	0.241	0.274	0.069	0.003
V114	0.105	0.174	-0.007	0.395	0.043	0.145
V116	0.192	0.153	0.272	0.110	0.065	-0.014
V118	-0.044	0.176	0.173	0.166	0.096	-0.018
V120	0.043	0.444	-0.036	0.378	-0.035	0.025
V122	-0.056	-0.041	0.092	0.471	0.010	-0.011
V124	-0.112	0.050	-0.075	0.487	0.013	0.017
V126	0.009	-0.044	0.041	0.290	-0.003	0.073
V128	-0.027	0.314	-0.152	0.162	0.023	0.179
V130	0.101	-0.054	0.268	0.216	0.147	-0.066

Table A16. Seven-Factor Solution for Vocabulary

	F1	F2	F3	F4	F5	F6	F7
V102	0.006	0.066	-0.063	0.389	-0.020	0.170	-0.082
V104	0.160	-0.016	0.489	0.012	-0.075	-0.075	-0.168
V106	0.038	-0.053	0.088	0.440	0.082	-0.020	-0.043
V108	0.053	-0.007	0.093	0.488	0.050	0.005	0.048
V110	-0.099	0.316	0.030	0.342	0.046	-0.043	-0.101
V112	-0.046	0.057	0.242	0.249	0.090	0.011	0.015
V114	0.087	0.243	0.004	0.336	0.017	0.120	-0.090
V116	0.167	0.171	0.284	0.075	0.028	-0.021	-0.065
V118	-0.088	0.226	0.207	0.110	0.036	-0.033	-0.083
V120	0.015	0.527	-0.005	0.283	-0.046	0.010	-0.033
V122	-0.033	-0.004	0.077	0.459	0.045	0.004	0.020
V124	-0.086	0.098	-0.088	0.464	0.049	0.032	0.028
V126	0.007	-0.012	0.038	0.279	0.002	0.072	-0.030
V128	-0.035	0.360	-0.134	0.104	0.014	0.160	-0.032
V130	0.043	-0.009	0.292	0.190	0.048	-0.083	-0.161

Table A17. One-, Two-, and Three-Factor Solution for Reading

	F1	F1	F2	F1	F2	F3
R132	0.328	0.276	0.115	0.050	-0.376	-0.044
R134	0.421	0.314	0.184	0.123	-0.318	0.084
R136	0.373	0.260	0.186	0.076	-0.286	0.108
R138	0.348	0.198	0.225	-0.040	-0.406	0.039
R140	0.538	0.411	0.222	0.137	-0.450	0.081
R142	0.498	0.406	0.180	0.144	-0.444	0.035
R144	0.391	0.305	0.167	0.078	-0.385	0.028
R146	0.466	0.290	0.277	-0.036	-0.530	0.041
R148	0.619	0.326	0.431	-0.063	-0.627	0.164
R150	0.494	0.284	0.321	-0.033	-0.502	0.117

Table A18. Four- and Five-Factor Solution for Reading

	F1	F2	F3	F4	F1	F2	F3	F4	F5
R132	0.028	-0.045	0.419	-0.107	-0.011	0.17	-0.026	-0.111	0.352
R134	0.111	-0.060	0.339	0.029	0.079	0.151	-0.019	0.027	0.285
R136	0.077	0.035	0.377	0.028	0.037	0.125	-0.098	0.013	0.327
R138	-0.051	-0.086	0.388	0.007	-0.027	0.060	0.066	0.008	0.375
R140	0.124	-0.024	0.508	-0.006	0.131	0.071	-0.018	-0.007	0.484
R142	0.129	-0.019	0.494	-0.039	0.155	0.026	-0.004	-0.043	0.492
R144	0.065	-0.025	0.429	-0.037	0.088	0.024	0.010	-0.035	0.426
R146	-0.047	-0.057	0.551	-0.019	-0.003	-0.004	0.067	-0.007	0.560
R148	-0.074	0.040	0.765	0.040	-0.040	-0.021	-0.019	0.049	0.766
R150	-0.038	0.004	0.583	0.027	-0.032	0.056	-0.025	0.033	0.558

Table A19. Six-Factor Solution for Reading

	F1	F2	F3	F4	F5	F6
R132	0.061	-0.089	-0.051	-0.044	0.296	0.202
R134	0.081	0.044	0.024	0.014	0.264	0.150
R136	0.073	0.033	-0.077	0.013	0.295	0.145
R138	0.096	-0.170	-0.047	0.124	0.288	0.126
R140	0.235	-0.081	-0.082	0.070	0.400	0.134
R142	0.169	-0.007	0.015	-0.027	0.468	0.036
R144	0.137	-0.060	-0.012	0.008	0.384	0.051
R146	0.010	-0.048	0.058	0.015	0.547	0.001
R148	-0.054	0.039	0.020	0.015	0.785	-0.030
R150	-0.072	0.081	0.051	-0.023	0.584	0.026

Table A20. Seven-Factor Solution for Reading

	F1	F2	F3	F4	F5	F6	F7
R132	-0.007	-0.059	-0.026	-0.045	0.109	0.157	-0.296
R134	0.017	0.083	0.057	-0.014	0.102	0.107	-0.256
R136	-0.049	0.105	-0.018	-0.031	0.044	0.080	-0.387
R138	0.001	-0.110	-0.013	0.115	0.067	0.073	-0.359
R140	0.054	0.009	-0.011	0.035	0.020	0.052	-0.583
R142	0.003	0.066	0.098	-0.065	0.120	-0.041	-0.512
R144	0.015	-0.001	0.043	-0.016	0.119	-0.008	-0.402
R146	-0.006	-0.051	0.055	0.020	0.421	-0.002	-0.186
R148	0.012	-0.005	-0.018	0.014	0.815	-0.021	-0.017
R150	0.002	0.034	0.012	-0.012	0.587	0.059	0.022

Table A21. Factor Correlation Matrices for Each Factor Solution

	F1	F2	F3	F4	F5	F6
Two-factor solution						
F2	0.381					
Three-factor solution						
F2	-0.429					
F3	0.171	-0.315				
Four-factor solution						
F2	-0.283					
F3	0.386	-0.340				
F4	0.207	-0.107	0.371			
Five-factor solution						
F2	0.485					
F3	0.142	0.225				
F4	0.184	0.277	-0.033			
F5	0.373	0.452	0.159	0.330		
Six-factor solution						
F2	0.225					
F3	0.373	-0.053				
F4	0.073	0.293	0.135			
F5	0.366	0.153	0.261	0.324		
F6	0.405	0.207	0.372	0.229	0.468	
Seven-factor solution						
F2	0.248					
F3	0.399	-0.023				
F4	-0.002	0.306	0.151			
F5	0.279	0.204	0.231	0.261		
F6	0.346	0.219	0.368	0.164	0.375	
F7	-0.417	-0.184	-0.299	-0.185	-0.599	-0.376

Appendix B

The following tables B1 and B2 present the parameter estimates of the factor loadings and item thresholds from the final five factor model. The noninvariant parameters are provided for each group separately.

Table B1. Invariant and Noninvariant Factor Loading Estimates by Subtest

	Low/Invariant	High
	Listening	
L2	1.488	0.926
L4	0.892	
L6	1.358	0.676
L8	1.649	1.066
L10	0.892	
L12	0.935	
L14	0.579	
L16	1.000	
L18	1.278	1.151
L20	1.327	0.998
L22	1.160	
L24	1.058	
L26	0.793	
L28	0.881	1.073
L30	1.278	0.859
L32	0.732	0.958
L34	0.754	
L36	1.394	0.804
L38	0.388	0.881
L40	0.774	0.858
L44	1.086	
L46	0.672	1.890
L48	0.579	1.510
L50	0.066	0.586
	Grammar	
G52	1.000	
G54	1.067	0.702
G56	0.603	0.909
G58	0.875	
G60	0.997	
G62	0.865	
G64	0.603	
G66	0.728	1.054
G68	1.141	0.870
G70	0.354	0.892
G72	1.085	1.210
G74	1.447	0.830
G76	0.910	0.655
G78	1.340	0.832
G80	1.700	0.924

	Low/Invariant	High
	Cloze	
C82	0.428	
C84	1.577	
C86	1.000	
C88	0.186	
C90	1.375	
C92	2.299	
C94	1.537	
C96	0.866	
C98	0.953	
C100	0.183	
	Vocabulary	
V102	1.077	
V104	1.011	
V106	1.000	
V108	1.164	0.966
V110	1.105	1.405
V112	0.580	1.068
V114	1.268	1.471
V116	0.927	1.151
V118	0.634	0.907
V120	1.101	1.642
V122	1.014	0.860
V124	0.806	
V126	0.662	
V128	0.372	1.212
V130	0.931	
	Reading	
R132	0.657	
R134	0.647	0.834
R136	0.494	0.887
R138	0.718	
R140	1.106	
R142	1.013	
R144	0.777	
R146	0.963	
R148	1.306	
R150	1.000	

Table B2. Invariant and Noninvariant Item Threshold Estimates by Subtest

	Low/Invariant	High		Low/Invariant	High
	Listening			Cloze	
L2	-1.286	-1.455	C82	-1.068	
L4	-0.564		C84	-0.357	
L6	-0.648	-0.756	C86	-0.699	
L8	-0.755	-0.846	C88	0.118	
L10	-0.210		C90	-0.164	
L12	-0.267		C92	0.455	
L14	-0.184		C94	-0.821	
L16	-0.636		C96	-0.817	
L18	-0.192	-0.192	C98	0.904	
L20	-0.409	-0.440	C100	-0.097	
L22	-0.747			Vocabulary	
L24	-0.234		V102	-0.791	
L26	-0.069		V104	-1.655	
L28	-0.008	0.021	V106	-0.867	
L30	-1.591	-1.613	V108	-0.057	-0.065
L32	0.135	0.154	V110	-0.147	-0.023
L34	-1.404		V112	0.356	0.449
L36	-1.239	-1.405	V114	0.094	-0.045
L38	0.768	0.879	V116	-0.431	-0.548
L40	-0.296	-0.351	V118	0.325	0.342
L44	-0.124		V120	0.574	0.656
L46	0.858	1.088	V122	0.229	0.250
L48	0.112	0.276	V124	0.475	
L50	0.477	0.562	V126	-0.394	
	Grammar		V128	0.326	0.463
G52	0.104		V130	-0.116	
G54	-1.147	-1.207		Reading	
G56	-0.073	-0.012	R132	-1.104	
G58	-0.074		R134	-0.567	-0.651
G60	-0.596		R136	-0.289	-0.242
G62	-0.437		R138	-0.293	
G64	-0.097		R140	-0.938	
G66	0.235	0.302	R142	-1.044	
G68	-0.479	-0.501	R144	-0.708	
G70	0.687	0.887	R146	-0.564	
G72	-0.673	-0.782	R148	-0.119	
G74	-0.983	-1.062	R150	-0.002	
G76	-0.423	-0.510			
G78	-1.542	-1.583			
G80	-1.824	-1.939			