PRF RESEARCH GRANT INFORMATION FORM

Brian F. French
Project Director - first, middle initial, last

Assistant Professor
Professorial Rank

Educational Studies
Department

494-7295
frenchb@purdue.edu
Telephone
Email

Education
School/College

BRNG
Bldg
West Lafayette
Campus

Title of Project
The Comprehensive Inventory of Basic Skills-II: Item and Test level Invariance

This research study will involve the use of: (check all that apply or include Ref. #[s] if known)

PACUC [Animals] □ IRB [Humans] ✓

IBC [r-DNA, infectious agents, or unfixed human fluids, human tissues or human cell lines] □

Do you currently have an active PRF Grant □ yes ✓ no
If yes, when is it due to expire __________________________ Date

Not required, however, if you know the name of the Ph.D. graduate student you will be supporting:

Name of Student________________________________________

Student ID # __________________________________________

Graduate Index______________ (Based on ____________ accumulative hours completed)

Project Director Signature

12/11/06

Date

Endorsed:

Dean or Delegate Signature

Date
The Comprehensive Inventory of Basic Skills-II: Item and Test Level Invariance

Abstract
The Comprehensive Inventory of Basic Skills (CIBS) is an assessment instrument used to identify students’ current academic performance level and connect assessment with instructional planning. The CIBS-II complies with the Individuals with Disabilities Education Act, No Child Left Behind Act, and many state standards. To compliment the standard psychometric analyses that will occur with the new version of the CIBS (i.e., CIBS-II), a series of studies will be conducted to examine item and test level measurement invariance across various groups (e.g., sex, race/ethnicity). Specifically, multi-group confirmatory factor analytic procedures and differential item functioning procedures will be employed to provide evidence of measurement invariance at the test and item level, respectively. This evidence will assist to ensure educational decisions and groups comparisons based on CIBS-II scores are accurate and fair.

1. Is this a 2nd year extension or competitive renewal application? NO

2. If this is not a 2nd year extension, have you ever had a year long PRF before, and if so, when? NO
Statement and Significance of the Problem

Educational testing has received much legal attention in the United States. The outcome of *Brown v. Board of Education* in 1954 began the attempt to achieve equality in education. The court ruled that segregated school systems were to end; yet equality was not obtained. Students were placed into certain classes based on measures of ability or achievement, which some persons claimed to be unfair. Improper classroom placement was brought to the court’s attention in *Larry P. v. Riles* in California. The case centered on the misuse of intelligence test scores to place African American students in special education classes as the tests were biased against African Americans. This was the first case at the federal level to demand validity evidence of test scores used for class placement (Wigdor & Garner, 1982). In the 1980 case of *Parents in Action on Special Education v. Hannon* a judged ruled, based on his subjective and non-expert opinion of item bias, that certain intelligence tests were free of cultural bias and therefore could be used for special education class placement of African American students (Wigdor & Garner, 1982). This type of evidence is not sufficient to claim that the test is unbiased (Plake, 1980). Legal cases will wax as testing is integrated into the educational system due to government requirements. For instance, parents in Texas claimed that the Texas Assessment of Academic Skills, a high school required exit exam, is biased against Latino and African American students (Saucedo, 2000). The issue of proper use of tests is a hotly debated topic not likely to lose the public’s interest any time soon.

The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) provide guidelines for all facets of testing (e.g., statistical properties, how results inform public policy). Standard 7.12 directly addresses the ethical issues involved in fair and unbiased testing by requiring that all persons tested obtain comparable and equitable treatment throughout the entire process. This includes equality in the inferences and decisions based on test scores.

An essential component in providing such evidence and ensuring equal treatment is the examination of measurement invariance. Measurement invariance (MI) is a necessary assumption for accurate group comparisons on a common latent trait using techniques such as latent means modeling. The violation of this assumption can lead to inaccurate comparative results on means of the latent trait, as
well as an incorrect understanding of the nature of the latent trait for the groups in question (Brown, 2006). Prior to comparing group differences, the psychometric properties of an instrument should be equivalent (e.g., configural, metric invariance; See Bollen, 1989; Meredith, 1993). Otherwise, interpretation of mean differences is questionable (Rock, Werts, & Flaugher, 1978), as statistically significant results could be due to real trait differences or to measurement problems.

Two methods used to detect measurement invariance are differential item functioning and multi-group confirmatory factor analysis (MCFA). DIF exists when persons from different groups (e.g., boys and girls) of equal ability (e.g., achievement, intelligence) have different probabilities of responding correctly to an item. MCFA allows for testing an a priori theory of the test structure across groups (McGaw & Jöreskog, 1971) or time (e.g., developmentally related questions, Mantzicopoulos, French, & Maller, 2004) to ensure the ability of interest is measured the same across the said conditions. An examinee’s score should not depend on their group membership (i.e., construct irrelevant variance). When decisions are made for individuals in the absence of measurement invariance, the decision maker risks committing a serious error (Bollen, 1989).

This validity evidence is needed for the Brigance Comprehensive Inventory of Basic Skills-II (CIBS-II), a revised measure used to inform educational decisions (e.g., Individual Educational Program development) for students 5 to 13 years old. Employing MCFA and DIF analysis to provide evidence for the CIBS-II scores across subgroups assists to satisfy minimum professional recommendations (e.g., AERA, APA, NCME, 1999), and thus provide a situation conducive to appropriate test score interpretation and decision-making in the educational environment.

**Proposed Study: Design and Research Methodology**

**Participants**

The standardization sample is currently being collected and will be complete by Fall of 07. The nationally representative sample will contain approximately 2,000 examinees ranging from kindergarten through 6th grade and are being selected based on U.S. Department of Educational Center for Education
Statistics (2006). The ethnic representation will be approximately 60% Caucasian, 15% African American, 18% Hispanic, and 7% Other. Sex representation will be approximately equal.

Instrumentation

The CIBS-II is an instrument that has a broad purpose and a range of applications. As a criterion-referenced measure, the CIBS-II is designed to: (a) measure mastery of developmental and academic skills; (b) identify areas of strengths and weaknesses; (c) serve as a longitudinal indicator of student progress; and (d) assist in identifying goals and objectives for individual plans. As a norm-referenced measure, the CIBS-II is designed to: (a) meet state and federal assessment requirements for the identification of exceptional students for purposes of placement within services; (b) assess six areas of academic achievement (basic reading skills including phonetic analysis and phonemic awareness, reading comprehension, math calculation, math reasoning, written language, and listening comprehension); (c) assess information processing skills in reading, math and written language designated under the Individuals with Disabilities Education Act for the assessment of learning disabilities; and (d) screen students to determine whether additional testing is needed. The CIBS-II will contain new and revised items for grades K-6 and new normative information overcoming limitations (e.g., limited standardization sample) of its predecessor. Psychometric evidence cannot be provided at this time for the CIBS-II as it is currently under investigation.

Study 1

Research Question

Do the items comprising the CIBS-II display DIF between sex and ethnic subgroups?

Analysis

Logistic regression (LR) analysis will be used for DIF detection. In ordinary logistic regression (LR), predictors are used to model the probability of observing a given level of an item response. As outlined by Swaminathan and Rogers (1990) the logistic regression model for DIF detection is:

\[ p(u_i = 1 | \theta, g) = \frac{e^{\beta_u + \beta_i \theta + \beta_j g + \beta_f (\theta g)}}{1 + e^{\beta_u + \beta_i \theta + \beta_j g + \beta_f (\theta g)}} \]  \hspace{1cm} (1)
where
\[ p_i = \text{Probability of correctly responding to item } i \]
\[ \beta_0 = \text{Intercept} \]
\[ \beta_j = \text{Slope for model term } j \]
\[ \theta = \text{Ability level, typically the total score} \]
\[ g = \text{Group membership}. \]

Predictors include the total score as an ability measure, a grouping variable (e.g., gender), and the interaction between ability and group. An item is flagged for DIF if the latter two variables significantly improve the model beyond the total score alone. A significant group or interaction term signals uniform or nonuniform DIF, respectively. A two-step purification of the matching criterion (i.e., total score) will be used prior to final DIF analysis, as it can improve DIF detection (Zumbo, 1999) and follows recommendations (Holland & Thayer, 1988). To classify an item as exhibiting DIF, the chi-square difference test is used to compare models as variables are entered into the model. Additionally, an \( R^2 \) value associated with each step is examined as the effect size measure. A significant chi-square difference test between models (\( p < 0.01 \)) and \( R^2 \) difference of at least 0.130 was used to flag DIF items (Zumbo, 1999).

Study 2

Research Question

Is the factor structure of the CIBS-II invariant across ethnicity (Hispanic, African American, & Caucasian) and sex subgroups?

Analysis

Maximum likelihood estimation via LISREL 8.7 (Jöreskog & Sörbom, 2005) will be used to conduct multi-sample confirmatory factor analyses using the covariance matrix of scores across groups. Model fit will be evaluated by several criteria: chi-square significance test, comparative fit index (CFI), and the standardized root mean square residual (SRMR), following recommendations (Hu & Bentler, 1999).

The CFA model for the \( g \)th group, upon which MCFA is built, takes the form (Jöreskog & Sörbom, 1996):
\[ x_g = \Lambda_g \xi_g + \delta_g, \]  

where \( x \) is a vector of observed variables, \( \Lambda \) is a matrix of factor loadings relating the factor to the observed variables, \( \xi \) is a vector of underlying factors, and \( \delta \) is a vector of measurement errors. Testing for factor invariance involves the comparison of increasingly more restricted models by sequentially constraining the factor loadings, error variances, and factor variances and covariances (e.g., French & Mantzicopoulos, in press). A significant decline in fit, judged by the chi-square difference, of a more restrictive model indicates that the factor model differs across groups. With the presence of invariance of these parameters, invariance of the latent mean structures are examined (Hancock, 1997; Little, 1997) by first assessing the invariance of the measurement intercepts (i.e., a predicted subtest score; Hancock). With invariant intercepts, latent mean differences can accurately be compared. If not, latent mean are not compared, as expected subtest score differences may reflect true group differences or systematic measurement bias (Hancock). An alpha value of 0.01 will be used for significance testing.

**Implications, Results, Interpretation**

The measurement of ability serves an important role in society, especially with the use of test scores for high-stakes educational decisions. In fact, a sharp increase in the use of test scores for these decisions has been observed (Brennan, 2004), especially in relation to No Child Left Behind Act of 2001 (NCLB; PL. 107-110). Thus, the statistical properties of tests must meet current validity standards (AERA, APA, & NCME, 1999) to overcome both legal and technical challenges (Brennan, 2004). This investigation will provide evidence as to the extent the CIBS-II measures abilities similarly across the subgroups examined. Results will (a) allow practitioners to make informed decisions about using the CIBS-II, (b) highlight potential areas of improvement on the instrument at the item and test level, (c) align the CIBS-II with the Standards in terms of validity evidence, and (d) stimulate applied and methodological research related to measurement invariance testing.
The role and duties of the Graduate Research Assistant

The graduate research assistant is an advanced doctoral student in Educational Psychology in the area of Applied Measurement and Research Methodology. The student has interest in psychometrics, including issues of measurement invariance, test validity, equating, and instrument development. The student will (a) manage daily activities related to the project, (b) manage data including entry, screening, and cleaning, (c) assist in acquiring resources needed to conduct analyses, (d) write programs to run the analyses, (e) continue to stay current with the literature in the area, (f) assist with writing technical reports and articles related to the project, and (g) present results at professional conferences.
Note the sustainability or future directions of research and plans for additional external support.

This proposed project will lead to four main outcomes related to future (a) applied research, (b) methodological research, and (c) external support. First, the analysis being conducted on the major subcategories of groups is only a fraction of the measurement invariance studies that need to be conducted with the CIBS-II. There is interest in examining similar issues with groups based on disability status and socioeconomic status, for example. This type of work will require application of various methods (e.g., Multiple Causes Multiple Indicators) to account for some low incidence populations. This will allow the student to gain more applied experience with dealing with advanced techniques for examining measurement invariance. Second, as the proposed methods are applied, it is anticipated that methodological studies will develop. For instance, measurement invariance methods and accuracy of detection under various missing data situations is an important topic. The real data in the proposed study will be the foundation for (a) a series of simulation studies looking at such issues and (b) providing parameters to accurately model reality in such studies. Third, it is anticipated that the project outcomes in conjunction with other psychometric analyses being conducted during the project time-period will lead to external funding to examine other such instruments for similar questions. For example, this additional component, not originally included in the larger CIBS-II project with Curriculum Associates, Inc, is assisting with establishing a relationship that is anticipated to continue. These analyses were not part of the original project but should become standard for such work. This is an essential step in moving in that direction. Fourth, this series of studies will likely lead to substantive questions in measurement invariance research that can be used to pursue funding from other external sources.
State the proposed timeline and rationale for funding.

The proposed studies require advanced training in statistical methods, psychometrics, and computer software due to the technical nature and required CPU and programming time. Several programs (e.g., SAS, MPLUS, LISREL, EXCEL) will be used to complete analyses, as common stand-alone procedures in SAS, for example, will not automatically conduct all steps required. Additionally, there are more than eight subtests in the CIBS-II and each subtest will be examined separately. Financial support will enable the student to focus only on the topics related to the proposed research while working as an apprentice to the faculty member. Supply funds will be used for needed resources (e.g., duplication, electronic storage, books), and student travel to a national conference to present the research.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Proposed Work Plan</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study 1</strong></td>
<td></td>
</tr>
<tr>
<td>DIF analysis</td>
<td></td>
</tr>
<tr>
<td>Purification</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>DIF Detection Analysis</td>
<td>6 7 8 9 10 11 12</td>
</tr>
<tr>
<td>Writing Article</td>
<td></td>
</tr>
<tr>
<td>Literature Review</td>
<td></td>
</tr>
<tr>
<td>Methods</td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td></td>
</tr>
<tr>
<td>Discussion</td>
<td></td>
</tr>
<tr>
<td><strong>Study 2</strong></td>
<td></td>
</tr>
<tr>
<td>MCFA analysis</td>
<td></td>
</tr>
<tr>
<td>Writing Article</td>
<td></td>
</tr>
<tr>
<td>Literature Review</td>
<td></td>
</tr>
<tr>
<td>Methods</td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td></td>
</tr>
<tr>
<td>Discussion</td>
<td></td>
</tr>
</tbody>
</table>
References

American Educational Research Association, American Psychological Association, &
Association.


Research Report, NO 4. University of Iowa. Iowa City, IA.

Guilford Press.

French, B. F., & Mantzicopoulos, P. Y. (in press). An examination of the first/second-
grade form of the pictorial scale of perceived competence and social acceptance: Factor structure
and stability by grade and gender across groups of economically disadvantaged children. Journal
of School Psychology.

Hancock, G. R. (1997). Structural equation modeling methods of hypothesis testing of
latent variable means. Measurement and Evaluation in Counseling and Development, 30, 91-105

Haenszel procedure. In H. Holland & H. I. Braun (Eds.), Test validity (pp. 129-145). Hillsdale,
NJ: Erlbaum.

Hu, L-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis:

Scientific Software International.

data: Practical and theoretical issues. Multivariate Behavioral Research, 31, 53-76.

pictorial scale of perceived competence and social acceptance with two pre-elementary samples.

*Child Development, 75*, 1214-1228.


