

Investigating the Moderating Effects of School Accountability Policies on the Relationship between Teacher Practices and Student Outcomes Internationally

Leslie Hawley¹, Betty-Jean Usher-Tate², Carina McCormick¹ and Sara Gonzalez²

¹Nebraska Center for Research on Children, Youth, Families & Schools, Lincoln, United States; ²Buros Center for Testing, Lincoln, United States

e-mail: lhawley2@unl.edu

School climate can influence students' motivation to learn and achieve (Deal & Peterson 2009). An aspect of school climate that influences teachers and subsequently students are the evaluation practices. In particular, an emphasis on accountability (e.g., students' test performance) may lead teachers to fear the threat of consequences (Santiago & Benavides, 2009). Consequently, a climate focused on accountability may lead teachers to alter their teaching practices, potentially influencing students' motivation and achievement. The purpose of this study is to evaluate the degree to which schools' teacher accountability policies moderate the relationship between teacher practices and student outcomes (motivation and academic achievement). The 2009 Programme for International Student Assessment (PISA) will be used for this project. A three-level (i.e., L1: student; L2: school; L3: country) hierarchical model will be analyzed using the HPMIXED procedure in SAS version 9.3. We will conduct two analyses with separate dependent variables: 1) teachers' stimulation of reading habits scale and 2) reading performance assessment score (incorporating plausible values methodology). Both analyses will include measures of teacher practices as predictors and accountability practices as moderators, as well as additional student- and school-level predictors to account for the socio-economic context of students, schools, and countries, providing policy-relevant results. This proposal represents a work in-progress. Data have been acquired and the initial steps have begun for the analyses, so there is sufficient time to complete the analyses in advance of the summer conference. One of the benefits of using hierarchical modeling is the ability to separate between-group and within-group effects of evaluation policies across schools and countries. Consequently, this research will contribute to the substantive literature regarding evaluation policies with student achievement data across different organizational and cultural contexts and expand existing literature by comparing these relationships in many countries.

Measurement of Gains in Intelligence – An Example of a Study Design and a Method of Data Analysis

Anna Hawrot and Aleksandra Jasinska

Educational Research Institute, Warsaw, Poland

e-mail: a.hawrot@ibe.edu.pl

In order to be able to estimate absolute differences between two given measurements of any variable, it is necessary that they are expressed on a common scale. The easiest way to achieve that is to use the same test in subsequent measurements. However sometimes it is not possible, e.g. due to the fact that the test measures too limited range of the variable of interest. The aim of the paper is to present an example of a study design and a method of data analysis allowing to estimate students' intelligence gains in a situation similar to the aforementioned one. We used data collected during a longitudinal study carried out as part of the "Research project for the development of educational value-added models". Random sample of 150 Polish lower secondary schools (5249 students) took part in the study. Intelligence was measured twice: when students were in the first (Raven's Standard Progressive Matrices) and in the third grade (selected items from Raven's Standard and Advanced Progressive Matrices). The results of two measurements were equated with use of the nonequivalent groups with anchor test design

(NEAT) with internal anchor. The statistical model was modified to take account of administering the tests to the same sample of students. Differential Item Functioning of the anchor items was verified. 2PL IRT model was used to estimate students' intelligence levels. The model allowed us to equate the results of two intelligence measurements and to estimate intelligence gains between the first and the third grade. Mean gain in intelligence equals 0,6 of standard deviation of results in the first grade. Furthermore, the size of the gains differs between schools. The results prove utility of the chosen study design and method of data analysis in estimating intelligence gains among adolescents.

Proposing a Shortened Version of Scandura & Graen's (1984) Leader-Member-Exchange Scale: Information Function and Validity

Ana Hernandez Baeza and Vicente Gonzalez Roma

University of Valencia, Valencia, Spain

e-mail: Ana.Hernandez@uv.es

Leader-Member-Exchange (LMX) (Dansereau et al., 1975) is a prolific theory for organizational research. According to this theory leaders develop a differential exchange with each of their subordinates, and the quality of these LMXs influences subordinates' attitudes and behaviors at work. One of the most common instruments to measure LMX is Scandura and Graen's (1984) 7-item scale. Even if this scale is not large, researchers frequently need to maximize the efficiency of their efforts to gather data by collecting information about a large number of variables. This makes it necessary to use very short scales with similar psychometric properties than the original scales. The aim of the study was to propose a shortened version of Scandura & Graen's scale by choosing the items that keep most of the information of the original scale. Samejima's (1969) Graded Response Model (GRM) is fitted to obtain the corresponding information functions for the original and the shortened scale. In addition, the two versions are compared in terms of criterion validity, considering a number of relevant criteria (e.g. well-being, commitment and propensity to leave). We use a sample of 426 health center employees to fit the GRM to the scale and evaluate the information functions and concurrent validity of the two versions analyzed. Moreover, for 182 of the employees, we use longitudinal data to compare versions' predictive validity. The 7 items clearly differ in their information functions. From the results we propose a shortened version made of three items that keep most of the information along the latent continuum. Regarding concurrent and predictive validity, both the pattern and the significance of correlations are similar for the original and the shortened version. The shortened version of the LMX scale is a good alternative to the original one both in terms of the information provided and criterion validity.

Education Level and Gender Differences on Spanish WAIS-IV Performance

Ana Hernández Fernández, Frederique Vallar and Èrica Paradell Calbó

Pearson Clinical and Talent Assessment, Barcelona, Spain

e-mail: ana.hernandez2@pearson.com

Clinicians are frequently required to estimate an individual's premorbid functioning. Procedures for adjusting the Wechsler scales for demographic variables (e.g., educational level) may buffers the effects of clinical status in predicting premorbid intellectual and memory functioning. Objectives. This study examines gender and education level differences on WAIS-IV performance in a Spanish sample. A Spanish general population sample (N=1002) stratified by gender, education level, region and population density was assessed with the Spanish WAIS-IV (Wechsler, 2012). ANOVA analyses were conducted to examine performance differences