Cross-Language Test Equating

Richard W. Woodcock, Ph.D. and Ana Muñoz-Sandoval, Ph.D.

How can performance levels in different languages be compared? Here are the steps we used to equate four tests, each developed to measure one of four skills (vocabulary, analogies, reading, and writing) in one language (English), to four similar tests in a second language (Spanish) for the *Woodcock-Muñoz Language Surveys* (Chicago: Riverside, 1993).

Step 1. Develop a bank of English items for each of the four skill areas.

Step 2. Rasch analyze the responses of subjects (6,359, ages 2 to 95) to establish the scale underlying each of the four banks of English items. Discard misfitting items.

Step 3. Compute norm measure tables of typical English language performance at different educational levels, ages, etc.

Step 4. Select sets of items for the four published English tests from their respective banks. Compute a raw score-to-measure table for each test for field use.

Step 5. Identify a subset of items covering the full difficulty range from "easy" to "difficult" in each English test and translate into Spanish. Verify that each English item has a reasonably direct counterpart in Spanish, e.g., authority/autoridad. (Though translated from English items of known difficulty, the difficulty of the Spanish items is still unknown.)

Step 6. Develop item banks for the four Spanish tests. Imbed in each bank of Spanish items the set of equating items translated from English.

Step 7. Rasch analyze the responses of Spanish speakers from outside the USA (1,200) and Spanish speakers from within the USA (800, pre-school to university graduate). Establish item difficulties for each Spanish item bank.

Step 8. Plot the English and Spanish difficulties for equivalent items for each skill area. Do not expect perfect agreement between the two sets of calibrations, because there are language-related differences for specific item content. Evaluate outliers from the best fit line for exclusion from the set of equating items. (This step verifies success in constructing comparable English- and Spanish-language variables.)

Step 9. Select items for the published versions of the Spanish tests. Compute a raw score-to-Spanish measure table for each test.

Step 10. Compute means (M_e for English and M_s for Spanish) and standard deviations (S_e for English and S_s for Spanish) of the English and Spanish equating item calibrations for each pair of tests.

Step 11. Equate each Spanish measure, B_s, to the scale of English measures, with value B_e.

Step 12. Construct a raw score to English equated measure table for each Spanish test, so that an examiner using the Spanish tests on native Spanish-speaking subjects can use the English equated measures to identify in the English language norm tables the typical performance levels for native English speakers of equivalent language competency.

Richard W. Woodcock
University of Virginia

Ana Muñoz-Sandoval
Measurement/Learning/Consultants
Box 161
Tolovana Park, OR 97145

■■■■■■■■■■■■■■■■■■■■

"Individual-centered statistical techniques require models in which each individual is characterized separately and from which, given adequate data, the individual parameters can be estimated. It is further essential that comparisons between individuals become independent of which particular instruments — tests or items or other stimuli — within the class considered have been used. Symmetrically, it ought to be possible to compare stimuli belonging to the same class — 'measuring the same thing' – independent of which particular individuals within a class considered were instrumental for the comparison."

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests* (reprint, with Foreword and Afterword by B. D. Wright) Chicago: University of Chicago Press, 1980, p. xx.

■■■■■■■■■■■■■■■■■■■■

T E S T I N G - T E S T I N G - T E S T I N G