

Title: Bateria Woodcock-Munoz--Revisada., By: Woodcock, Richard W., Munoz Sandoval, Ana F., Vol. 13

Database: Mental Measurements Yearbook

Bateria Woodcock-Munoz--Revisada.

Review of the Bateria Woodcock-Munoz--Revisada by ROBERT B. FRARY, Professor and Director Emeritus, Office of Measurement and Research Services, Virginia Polytechnic Institute and State University, Blacksburg, VA:

The Bateria Woodcock-Munoz: Pruebas de Habilidad Cognitiva--Revisada (Bateria--R COG) and the Bateria Woodcock-Munoz: Pruebas de Aprovechamiento (Bateria--R APR) expand and supplant the Bateria Woodcock Psico-Educativa en Espanol (Woodcock, 1982). The two new batteries are designed as parallel Spanish-language versions of the Woodcock-Johnson Tests of Cognitive Ability--Revised (WJ-R COG; 1989) and the Woodcock-Johnson Tests of Achievement--Revised (WJ-R ACH, 1989). As such, they share the complex structure of the Woodcock-Johnson (12:415) batteries, which comprise 35 separate tests providing norm-referenced scores, combinations of which yield a number of additional scores. Any evaluation of the Bateria-R batteries is in part dependent on the qualities of the WJ batteries. In this regard, the reader should consult reviews by Webster (1994), Cummings (1995), and Lee and Stefany (1995). These reviews (especially the review by Webster) describe the characteristics of the WJ batteries and the theory underlying their development in some detail. Because this information applies as well to the Bateria-R batteries, what follows is only an overview.

Like their parent batteries, the Bateria-R COG and APR provide a comprehensive set of individually administered tests of abilities and achievement spanning a wide range of ages (2 to 90). They are not intended to be administered in their entirety but selectively according to the needs of the individual examinees. Indeed, it might take 2 to 3 days to administer all 35 tests in both batteries to an individual. Each battery is divided into a 'standard' set of tests and a 'supplementary' set. Combinations of tests from within these groupings yield up to 30 additional scores. For example, the Bateria-R COG tests for Memoria para Frases (Memory for Sentences), Pareo Visual (Visual Matching), Integración de Sonidos (Sound Blending), and Vocabulario Oral (Oral Vocabulary) yield a 'cluster' score, Aptitud en Lectura (Reading Aptitude). The development of the WJ-R COG (and hence the Bateria-R COG) battery was based on the Horn-Cattell model of intellectual processing (Horn & Cattell, 1966). There are seven standard Bateria-R COG tests, each of which represents one of the abilities delineated by Horn and

Cattell. The WJ-R ACH (and hence the Bateria-R APR) battery is organized from a theoretical standpoint that distinguishes tasks according to their complexity and whether they require connected discourse or unitary responses. Again, combinations of tests yield cluster scores. For example, the Science, Social Studies, and Humanities tests yield a Broad Knowledge score.

The following features distinguish the Bateria-R from the WJ-R:

1. A brief preliminary Language Use Survey collects information useful in interpreting testing outcomes.
2. A Comparative Language Index may be determined for examinees who take both the Bateria-R and the WJ-R. This index not only indicates the relative strength of the examinee in Spanish versus English but also indicates a norm-based level of proficiency in each language.
3. The Bateria-R COG cluster score, Oral Language, may be transformed into the five-level scale of Cognitive-Academic Language Proficiency (Cummins, 1984).
4. A Supplemental Manual accompanies each Bateria-R battery. The features just listed are explained extensively in these manuals.

The Supplemental Manuals refer to a computer program, Report Writer for the Bateria-R, which has not been produced. Instead, capability to handle data from the Bateria-R has been incorporated into the Woodcock Scoring and Interpretative Program (Schrack & Woodcock, 1997), which also handles data from the WJ-R (B. Wendling, personal communication, September 6, 1997). It generates score reports and narrative interpretative reports including comparative language information. These reports are in English except for a summary report available in both languages. This program is available from the publisher (\$395) separately from either of the Bateria-R (or WJ-R) batteries. It operates on a Windows(TM) platform; a Macintosh(TM) version is anticipated.

Administration of the Bateria-R is essentially the same as for the WJ-R, which requires substantial experience in the testing of individual subjects and strong familiarity with the

administration and scoring procedures for each of the 35 tests. Moreover, for the Bateria-R, the examiner must be fluent in Spanish. Because need for the Bateria-R may occur in settings where a fully qualified Spanish-speaking examiner is not available, each Supplemental Manual provides comprehensive instructions for the training and utilization of ancillary Spanish-speaking examiners. The manuals make a sharp contrast between simply using such a person as an interpreter as opposed to requiring more active involvement with the measurement process.

It was the consensus of the reviewers cited earlier that the WJ-R embodied a high level of quality and that it had the potential to yield reliable and valid scores when properly used for its intended purposes, such as provisional placement in academic programs and diagnosis of learning deficiencies. They viewed the norming of the WJ-R as an especially strong point, given the size and comprehensiveness of the norming sample. The extent to which these qualities extend to the Bateria-R constitutes the focus of this review.

A complex test development and equating procedure was undertaken as described in each Supplemental Manual. For each WJ-R test, equating items were identified. These were items that could be adopted directly (nonverbal or mathematical items) or translated unambiguously into Spanish and which spanned a wide range of difficulty based on Rasch model calibration of the responses of the WJ-R norming sample. Additional items were written in Spanish, both to cover English items that resisted translation and to produce additional Spanish items for strictly verbal tests. The resulting Spanish item bank was administered in several countries, including the U.S., to nearly 4,000 monolingual subjects from 2 years of age through university graduate students (age unspecified). The equating procedures permitted creating raw-score to Rasch (W) ability tables consistent with the U.S. norms. Thus, a raw score on a Bateria-R test can be evaluated for percentile rank, grade-equivalence, etc., by entering the appropriate U.S. norms table with the corresponding transformed Rasch ability measure. A more detailed description of this process is found in Woodcock and Munoz-Sandoval (1993).

Though the development/equating process would probably be endorsed by most measurement specialists, its success must be judged in terms of the tests themselves and the psychometric characteristics of their scores. One of the most important concerns is content validity, especially given the possibility of content changes associated with translation. Many Bateria-R tests are virtually or largely identical to those of the WJ-R. These are the nonverbal and mathematics-based tests, for which only the instructions and brief item wordings had to be translated. The Spanish translations or adaptations for these tests appear to be entirely adequate. Tests with more extensive verbal stimuli contain many items completely different from the WJ-R. In some cases, these changes were intentional so that administration of a Bateria-R test would not give away the

content of the corresponding WJ-R test for bilingual examinees taking both. In other cases, the changes reflect adaptation to Spanish language or culture. The new items and items that underwent fairly extensive changes in translation employ standard Spanish in a manner that should be readily understandable across a wide variety of backgrounds for speakers of this language. Moreover, with some exceptions to be noted, the content, tone, and character of the translated items is very similar to those of the WJ-R that they replace.

One pair of tests that clearly measure different constructs are the WJ-R Letter-Word Identification test as contrasted with Identificación de Letras y Palabras in the Bateria-R APR. The latter items of both of these tests require the subject to read and then state orally words of increasing complexity. In English, this task requires knowledge of a highly complex and illogical set of orthographic/pronunciation rules and exceptions, to say nothing of the need to know how specific words are pronounced. In Spanish, however, one needs only to know a small set of consistent rules. Memory and vocabulary level must be major determinants of success for the English test but not for the Spanish. Therefore, it would be inappropriate to say that Anglophone and Hispanophone subjects attaining the same percentile rank on these tests are somehow equivalently endowed with respect to the task at hand. Nevertheless, the differing demands of this task in English and Spanish may not be a serious impediment to interpreting the scores in a clinical setting. In contrast, if the results were to be used comparatively, perhaps for program evaluation, the potential for misinterpretation is much greater.

It would have been reasonable to expect content/construct differences between the corresponding Social Studies tests of the Bateria-R and WJ-R, inasmuch as the latter has some orientation toward U.S. institutions and history. For example, the WJ-R contains questions requiring knowledge of why the Pony Express was discontinued and the name of the branch of the U.S. federal government responsible for law enforcement. Contrary to what might have been expected, similar questions are found in the Bateria-R APR test, Estudios Sociales, in which examinees are asked to identify the work of George Washington Carver and name the country that gave the Statue of Liberty to the U.S. Accordingly, examiners need to be aware that the Bateria-R Social Studies score represents to some extent achievement in social studies as they might be taught or presented in the U.S., not what may have been learned in a Spanish-speaking country. However, the proportion of such U.S.-oriented questions is relatively small in either test and should not influence scores to an extreme extent.

In contrast to the amount of U.S.-dependent content in the Bateria-R Social Studies test, there is essentially none in the Bateria-R Humanities test. There are only a few U.S.-dependent questions

in the WJ-R Humanities test (e.g., concerning the theme of *The Yearling* and the author of Tom Sawyer), and these or similar ones do not appear in the corresponding Bateria-R test.

Another problem associated with cross-national interpretation of testing outcomes is revealed in the work of Fletcher and Sabers (1995), who analyzed norming data from several countries for the Bateria Woodcock Psico-Educativa en Espanol (Woodcock, 1982), which contains tests highly similar to some of the Bateria-R. Fletcher and Sabers found that means based on cluster scores from the Bateria Woodcock varied substantially and quite irregularly across countries and grade levels. They conjectured that this outcome was due to differences among educational systems. Again, clinical judgment should prevent serious misinterpretation of the scores of an individual. However, knowledge of intercountry educational differences may be more than can be expected of many clinicians.

Perusal of the items of the various tests revealed a moderate number of item-writing or other editing lapses. For the most part, these problems stem from overly restrictive decisions about correct answers or problems with translations of words or picture meanings from English to Spanish. These oversights need not be considered flaws, especially if clinicians are sufficiently knowledgeable and flexible in scoring. The publisher should be sensitive to these issues in future printings.

The Supplemental Manuals provide information about reliability and validity. Internal consistency reliability coefficients reported are generally good. However, these coefficients are reported for only 9 of the 21 tests in the Bateria-R COG, whereas all tests in the Bateria-R APR are covered. Validity evidence reported in the Supplemental Manuals is meager. An unpublished study involving 70 U.S. kindergarten students found substantial correlations between scores from the Oral Language cluster of the Bateria-R COG and other measures of oral fluency in Spanish. Another unpublished study of 120 U.S. second grade students found moderate correlations, again between Bateria-R Oral Language scores and other measures of oral language.

SUMMARY. In spite of very limited empirical evidence of the adequacy of the Bateria-R COG and APR and the flaws noted above, there are, nevertheless, considerable grounds for recommending these batteries. They are the only thoroughly comprehensive ones suitable for evaluation of Spanish-speaking students in U.S. schools and could well be used in other English-speaking educational settings. Tying score outcomes to U.S. norms argues to some extent against their use in strictly Hispanic settings as does the use of U.S. money, U.S. time zones, U.S. ZIP

codes, and the English system of measurement in a number of items. The testing materials are well organized and of good physical quality. Most important, profiting from experience with the WJ-R, the authors have produced tests that should be substantially functional for use with their intended population.

REVIEWER'S REFERENCES

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57, 253-270.

Woodcock, R. W. (1982). *Bateria Woodcock Psico-Educativa en Espanol*. Itasca, IL: The Riverside Publishing Co.

Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and psychology*. San Diego: College Hill Press.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Tests of Achievement--Revised*. Itasca, IL: The Riverside Publishing Co.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Tests of Cognitive Ability--Revised*. Itasca, IL: The Riverside Publishing Co.

Woodcock, R. W., & Munoz-Sandoval, A. F. (1993). An IRT approach to cross-language test equating. *European Journal of Psychological Assessment*, 9, 233-241.

Webster, R. E. (1994). Review of Woodcock-Johnson Psychoeducational Test Battery--Revised. In D. J. Keyser & R. C. Sweetland (Eds.), *Test critiques* (vol. X, pp. 804-815). Austin, TX: PRO-ED, Inc.

Cummings, J. A. (1995). [Review of the Woodcock-Johnson Psychoeducational Battery--Revised.] In J. C. Conoley & J. C. Impara (Eds.), *The twelfth mental measurements yearbook* (pp. 1111-1116). Lincoln, NE: The Buros Institute of Mental Measurements.

Fletcher, T. V., & Sabers, D. L. (1995). Interaction effects in cross-national studies of achievement. *Comparative Education Review*, 39, 455-467.

Lee, S. W., & Stefany, E. F. (1995). [Review of the Woodcock-Johnson Psychoeducational Battery--Revised.] In J. C. Conoley & J. C. Impara (Eds.), *The twelfth mental measurements yearbook* (pp. 1116-1117). Lincoln, NE: The Buros Institute of Mental Measurements.

Schrank, F. A., & Woodcock, R. W. (1997). *Woodcock scoring and interpretive program*. Itasca, IL: The Riverside Publishing Co.

Review of the Bateria Woodcock Psico-Educativa en Espanol by MARIA PRENDES LINTEL, Assessment Coordinator, Lincoln Family Practice Program, Lincoln, NE:

[Editor's Note: This review is based on the Bateria Woodcock Psico-Educativa en Espanol, which is the earlier edition of the Bateria Woodcock-Munoz described above.]

The Bateria Woodcock Psico-Educativa en Espanol (Bateria) was developed in 1982 and according to the manual has been widely used since its development. However, it is extremely difficult to evaluate this test, which purports to have an extensive variety of applications in educational and noneducational settings. The reason for this is that the manual does not discuss validity of any type or reliability factors but does state that a future publication of the technical manual is planned. However, that was in 1982 and in 1996 the manual was not among the components available. There are other concerns aside from those previously listed. One major concern is the norming sample, the other is the translation of the Bateria to Spanish.

NORMS. The norms for the Bateria, according to the manual, are based on a composite sample gathered during January to July of 1980 from typical urbanized areas in Costa Rica (San Jose), Mexico (Chihuahua and Guadalajara), Peru (Lima), Puerto Rico (Ponce), and Spain (Madrid).

There were 802 subjects in the sample drawn from kindergarten, grade 1, grade 3, grade 5, grade 8 (2 secundaria), and grade 11 (2 bachillerato, 3 BUP). The manual indicated that 'General Spanish Norms provide functional and instructional information regarding how well a subject is faring compared with native Spanish speakers across the Spanish speaking world' (manual, Woodcock, 1982, p. 28). The Bateria is said to have a common norm base, meaning the same group of people provided all the normative data for the Bateria. The above statement suggests scores from this test would be valid for anyone in the Spanish-speaking world regardless of linguistic bias, cultural uniqueness, history, and conceptual differences. Such statements are of concern and are perhaps reflective of the times in which this test was developed. However, given the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1985) and the American Psychological Association's (1990) Guidelines for Providers of Services Ethnic, Linguistic, and Culturally Diverse Populations, testing psychologists should not assume the appropriateness of results from using this test.

I will briefly summarize the reasons why the Bateria should not be used. First, the manual does not discuss how the test was developed or translated into Spanish. There are issues of conceptual meaning and understanding as well as ethnic and linguistic bias that are totally absent from the manual. Second, it is inappropriate to use a test normed on a population other than the one being tested (e.g., normed on Spanish children from other countries) as such norms are not representative of Hispanic children in the United States. Third, the manual does not address nor present any validity or reliability information. A better practice for testing the Hispanic population in the United States includes but is not limited to: (a) A review of test items for appropriateness of content, editorial accuracy, and ethnic and linguistic bias; (b) the test should be normed on a representative sample of the various Hispanic minorities population in the United States, including geographic representation; (c) there should be a test manual where appropriate levels of reliability, validity, and statistical analysis are fully described and reported.

It is my understanding that a new version of the Bateria will soon be introduced. Because I have not had the opportunity to review the new version I can only hope that it is not like its predecessor. The Bateria is questionable for use even in those countries in which it was normed because it is still missing vital information (e.g., reliability, validity) necessary for a testing psychologists to have in order to make a professionally responsible decision regarding the use of this test. The Bateria test developers are referred to the Spanish Assessment of Basic Education, Second Edition (291) for a model of appropriate test development and measurement with the Hispanic population of the United States.

REVIEWER'S REFERENCES

Woodcock, R. W. (1982). Examiner's manual Bateria Woodcock Psico-Educativa En Espanol, Woodcock Spanish Psycho-Educational Battery. Itasca, IL: The Riverside Publishing Co.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). Standards for educational and psychological testing. Washington, DC: APA, Inc.

American Psychological Association. (1990). Guidelines for providers of services ethnic, linguistic, and culturally diverse populations. Washington, DC: Author.