Annotations on the Use of the Mexican Norms for the WAIS-III

Authors: Pedro Sánchez Escobedo [a]; Liz Hollingworth [b]
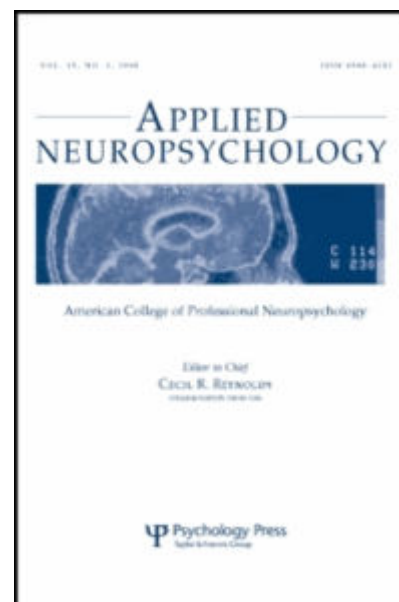
Affiliations: [a] Facultad de Educación, Universidad Autónoma de Yucatán Mérida, Yucatán, México
[b] Educational Policy and Leadership Studies, The University of Iowa, Iowa City, Iowa

Abstract

This article provides crucial information to judge the appropriateness of the Mexican version of the Wechsler Adult Intelligence Scale-Third Edition and recognizes some limitations in both the process of its adaptation to the Mexican population and the norm development process. This is an effort to contribute to the debate initiated by Suen and Greenspan (2008), who argued in court against the use of Mexican norms in a death penalty case, which depended upon establishing the diagnosis of mental retardation. As a part of the defense team, these scholars argued a number of points against the use of the Mexican norms. With input from the lead researcher on the Mexican standardization process, some of the criticisms are addressed, and further information about the norm development process for this test in Mexico is provided in an attempt to be critical about the strengths and weaknesses of the use of existing Mexican norms. Finally, we argue that results from a single test must not be used to make life and death decisions and that test development is a continuous process influenced by culture, language, and indeed by norm-developing procedures and debates.

Keywords: cross-cultural testing; intelligence testing; WAIS-III

Since 2001, efforts have been made to adapt and develop norms for the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III) in Mexico. The basic tenet of this process is the belief that psychometric tests are influenced by culture, language, and more generally, the social conditions which cause these effects (Weiss, **2003**, p. 50). Thus, even well-accepted, widely used assessment batteries in the United States should undergo a process of revision when used with international populations.

Using tests in different cultural contexts requires some judgment. In most cases, items need to be carefully translated and worded, considering different accepted words and phrases that are synonymous for the same concept and also taking into account the uses of various terms in a given population. Furthermore, even in standard translations, some items need to be adapted, and its criteria for acceptance revised. For example, consider the case of the translation of a vocabulary item: "What is a suburb?" as "¿Qué es una colonia?" In Spanish, test developers must consider the possibility of a respondent answering, "Cologne, fragrance, or perfume" as a correct answer in Spanish. Without a doubt, adaptation of a test to

another language requires more considerations than the mere translation of items, such as uses of the terms, differential meanings, and alterative definitions.

The purpose of this article is to address various critiques of the standardization of the WAIS-III in Mexico, discussed by Suen and Greenspan (**2008**), hired by the defense of a 37-year-old Mexican man accused of murder and for whom the defense attempted an Atkins exemption of the death penalty due to mental disability. In this case, when using American norms, the defendant met the criteria to suspect a borderline mental retardation (MR), but when he was assessed with the Mexican norms, his intelligence scores could be interpreted as above the threshold for MR. As a result, the defendant was at risk of being sentenced to death if the Mexican norms were used to interpret his score. Although, in this article we do not address the peculiarities of this very illustrative case, we do argue about the uses of norms and psychometric tests in general and on the difficulties and barriers in adapting such tests to other populations.

The perspective of academics involved in the Mexican standardization process is necessary to understand the limitations of adapting American batteries to assess Mexicans and also to illustrate the advantages of using Mexican norms to assess Mexicans. This article also attempts to supplement the information contained in the technical manuals of the Mexican standardization process (Tulsky & Zhu, **2001**, **2003**) and aims to recognize the difficulties and limitations in developing norms for American psychometric batteries abroad.

THE NORM DEVELOPMENT PROCESS

The process for developing norms for any large-scale test is iterative. That is, periodically and after a significant administration of the tests, the norms are revisited and the scoring scales are adjusted. In particular, the first attempt to use a test with a different population than the one initially used to establish the original norms requires that the norms are continuously revised and adjusted to the effects of time, practice, etc. with the new population.

In the case of Mexico, in 2001, a preliminary version was developed to adjust norms solely for the group of reference. Then, a more extensive and rigorous standardization process was carried out in 2003 to establish norms by age group. In 2004, new revised norms were published to adjust some discrepancies reported by users. Specifically, a statistical correction was made for the underestimation of IQ in the Mexican population when both American and Mexican norms were used. The acknowledgment of this process is crucial, since Suen and Greenspan failed to recognize the existence of different versions and times. In addition, it is unclear which version of the WAIS-III was administered to the defendant in the Atkins case—the Hispanic version of the American test (1999), the preliminary Mexican version (2001) or the latest version in 2003—and whether their critiques are stemming from the norms from 2001, 2003, or the 2004 Mexican version.

In order to be thorough in addressing Suen and Greespan's more substantive concerns, the following sections discuss some of the most important points under debate, attempting to provide information about what was not included in the published technical manuals.

THE RELIABILITY ISSUE

Reliability scores were not published in the manual. For this revision, we have calculated from the original database the alpha scores for each age group. Table **1** depicts the calculated coefficients from the 2001 administration. Notice that the majority of these coefficients are within the range of what is usually acceptable. These were obtained from the experimental forms. So far, no information is available about the reliability of the latest and currently used version of the test (2003) that includes changes and revisions from the standardization procedure.

TABLE 1 Reliability Scores (Cronbach's Alpha) by Age Group

Age group	16-17	18-19	20-24	25-29	30-34	35-44	45-54	55-64	65-69	>70
Vocabulary	0.80	0.77	0.87	0.77	0.83	0.53	0.86	0.87	0.90	0.32
Similarities	0.71	0.79	0.68	0.86	0.83	0.63	0.80	0.74	0.82	0.50
Arithmetic	0.86	0.71	0.75	0.77	0.76	0.80	0.42	0.72	0.80	0.86
Digit span	0.82	0.62	0.81	0.79	0.83	0.79	0.63	0.80	0.77	0.86
Information	0.79	0.85	0.80	0.82	0.36	0.90	0.87	0.28	0.83	0.82
Comprehension	0.71	0.69	0.88	0.87	0.84	0.66	0.87	0.89	0.83	0.72
Letter/Number	0.86	0.84	0.70	0.83	0.79	0.89	0.62	0.84	0.77	0.90
Picture completion	0.52	0.49	0.62	0.38	0.63	0.77	0.69	0.64	0.78	0.82
Block design	0.74	0.73	0.61	0.62	0.82	0.84	0.83	0.70	0.76	0.77
Matrix reasoning	0.69	0.70	0.25	0.72	0.78	0.79	0.90	0.85	0.79	0.87
Picture arrangement	0.77	0.74	0.56	0.75	0.79	0.71	0.84	0.76	0.55	0.72
Object assembly	0.50	0.60	0.52	0.47	0.46	0.56	0.61	0.64	0.70	0.74