

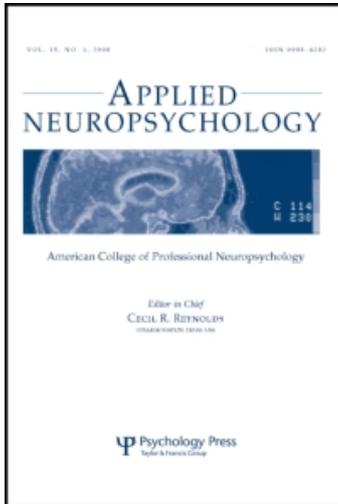
This article was downloaded by: [Hollingworth, Liz]

On: 5 August 2009

Access details: Access Details: [subscription number 913658177]

Publisher Psychology Press

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Applied Neuropsychology

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title-content=t775648089>

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

Pedro Sánchez Escobedo ^a; Liz Hollingworth ^b

^a Facultad de Educación, Universidad Autónoma de Yucatán Mérida, Yucatán, México ^b Educational Policy and Leadership Studies, The University of Iowa, Iowa City, Iowa

Online Publication Date: 01 July 2009

To cite this Article Escobedo, Pedro Sánchez and Hollingworth, Liz(2009)'Annotations on the Use of the Mexican Norms for the WAIS-III', Applied Neuropsychology, 16:3, 223 — 227

To link to this Article: DOI: 10.1080/09084280903098851

URL: <http://dx.doi.org/10.1080/09084280903098851>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

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

Pedro Sánchez Escobedo

Facultad de Educación, Universidad Autónoma de Yucatán Mérida, Yucatán, México

Liz Hollingworth

Educational Policy and Leadership Studies, The University of Iowa, Iowa City, Iowa

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.

Key words: 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 alternative 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.

Address correspondence to Liz Hollingworth, The University of Iowa, College of Education, 340C Lindquist Center, Iowa City, IA 52242. E-mail: liz-hollingworth@uiowa.edu

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 Greenspan'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.

REFERENCE POPULATION

We agree with Suen and Greenspan: IQ scores should be interpreted as to how well an individual relates to a

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

TABLE 2
Participants by Gender and Region

	North		West		Central		East		Total	
	<i>f</i>	%								
Men	114	48	102	47	107	44	125	47	410	42
Women	124	52	117	53	128	56	144	53	456	48
Total	238	25	219	23	244	25	269	27	970	100

specific population. Indeed, the Mexican technical manual recognizes these limitations and stresses the importance of considering the characteristics of the norming sample. Limited resources for the standardization in Mexico did not allow for a random stratified sample of the Mexican population, so developers attempted to obtain a sensible conventional sample by controlling, a priori, for some of the major variables affecting norms (see Table 2).

For instance, we intended to include half of the participants with low educational levels (<8th grade) and half who were better educated; half of the planned sample were to be women, and proportional numbers of participants were aimed by age group and by region of the country. Most importantly, clear exclusion criteria were depicted—the test was not to be administered to people with obvious physical or mental handicaps, people from rural towns, or people whose native language is not Spanish. For this reason, test users are urged to use the American norms for specific populations or to use caution with this version of the WAIS with specific populations, such as people with evident and specific disabilities (not necessarily with intellectual borderline patients who in most cases are not obvious to the tester). Therefore, the reference population was 970 Mexicans. Table 3 illustrates participants by gender and geographical zone.

TABLE 3
Participants by Age Group and Level of Education

Age	Level of Education		Total	%
	<i>H</i>	<i>L</i>		
16–17	53	46	99	10
18–19	78	32	110	12
20–24	62	33	95	10
25–29	64	33	97	10
30–34	53	37	90	9
35–44	54	43	97	10
45–54	50	44	94	10
55–64	55	43	98	10
65–69	32	45	77	8
>70	55	52	107	11
Total	556	408	964	100

Note. H = high level of education (above 9th grade); L = low level of education (below 9th grade).

Age groups were also different from American norms and were reduced to 10 bands, basically by clustering Mexicans 70 years old or more in one group, since Mexicans' life expectancy and proportions of population in elder groups are significantly lower than the American population (see www.inegi.gob.mx for extensive data on the Mexican population). Table 3 depicts participants by age group and level of education.

There are several advantages to this sampling strategy. For example, there were 52% women (consistent with the Mexican population), and a fairly balanced number of participants per age group was achieved. When the data from the sample were analyzed, a bias toward including better-educated Mexicans was discovered (57/43%, $\chi^2 = 4.86$; $p \leq .001$). Further exploration about the possible causes of this effect led the research team to suspect that criteria for participant selection were not followed for 3% of the sample. In many cases, hired assistants found it easier to administer the test to fellow students or family members, disregarding the instructions for participant recruitment. This is, in fact, the most tenable hypothesis of why in the first draft, the Mexican norms tended to underestimate the IQ, and this is why the norms were revised in 2004. The scores from the participants who were recruited in violation of the recruitment guidelines (3% of the sample) were thrown out of the sample, and the norms were recalculated.

In sum, sufficient information exists with reference to the population that can be tested (and not) with the Mexican version. Although, in theory, a random stratified sample is ideal in practice; conventional samples are commonly used in the social sciences to approach diverse issues.

SCORE NORMALIZATION METHODS

Raw scores were transformed to normalized scores. Suen and Greenspan jump to an unfortunate conclusion when considering the statistical calculations of the scale scores. Standardized scores were indeed calculated using the formula:

$$DE \times \left(\frac{PE - 10}{3} \right) + \mu$$

Thus, what is framed as a “glaring statistical error” in the review by Suen and Greenspan can easily be explained as a typographical error in the preliminary draft of the 2001 manual, not a technical error that would invalidate the use of the scales.

The method of linear standard scores was used to calibrate norms. Obtained normalized scores were compared in the 2003 stage to the mean and ± 1 standard deviation in order to estimate the degree of coincidence

between them, and graphical and plotting methods were used to calibrate norms in 2004.

LACK OF REPRESENTATION OF CERTAIN GROUPS

It is a recognized fact that the Mexican norms are not to be used with special populations, such as people with major disabilities, those who are incarcerated, adults with brain damage, dementias, etc. This was stated in the *WAIS-III Guide for Use and Interpretation* (The Psychological Corporation, 1997, p. 13; Sanchez, 2001, 2003), and it is sustained here. Thus, individuals with obvious MR, for instance, may be better assessed using the specialized American norms rather than the Mexican Norms. In short, the Mexican norms represent adults in urban areas with no special educational or mobility needs.

USE OF OTHER STATISTICS AND CALCULATIONS

Suen and Greenspan suggested a number of miscalculations and a poor use of statistical procedures and provide as an example “the lack of information about the degrees of freedom . . .” of routine statistical comparisons from which these elements could be simply inferred. We used a χ^2 statistic to compare participants from the low and high educational levels. Because the formula for degrees of freedom is $df = N - 1$, and $N = 3$, the $df = 2$.

In any case, the point was clearly made about the significant over-representation of better-educated Mexicans in the sample. We noticed, however, some mistakes in the published user’s manual, and we appreciate their feedback—or example, the typo in the formula to calculate standard scores, the misuse of \geq instead of \leq when establishing the probability level, and when we wrongly attributed less reliability to the smaller sample size.

Despite all these errors and faults in the publication of the norms and its manuals, we do argue that the Mexican norms are better suited to assess Mexicans, because they compare Mexicans against other Mexican adults.

INTELLECTUAL DISABILITY IN MEXICO

Finally, since this case emerged from a debate on the diagnosis of a case of MR or intellectual disability (ID), in which Suen and Greenspan claimed a lack of definition and boundary parameters of ID in Mexico, we think this is an outstanding real-life event to reflect upon psychometric testing and assessment in different cultural contexts.

Suen and Greenspan’s suggestion that the use of certain norms may be the difference between life and death decisions is misleading. Beyond legal and commercial issues, it of course would be unfair to elicit a high-stakes judgment based on the results of a single test. In fact, the American Association on Intellectual and Developmental Disabilities warns against establishing the diagnosis of MR based upon solely on IQ scores. A criterion of adaptive behavior is necessary. In the transcripts of the trial available to us, no discussion was present about the subject’s adaptive behavior and living skills in this particular case (apparently there was a discussion on this issue, which was not considered in the Suen and Greenspan article). No test result, not even with almost perfect norms, should be the sole criterion used to make important decisions about a person. Many other measures and additional qualitative information are necessary to fully understand a person and to place their test scores in context.

This is also a point of reflection: Although in Mexico scholars have stated both cognitive and adaptive behavior parameters to establish the diagnosis of MR and its degree (e.g., Sanchez, Canton, & Sevilla, 1999), no test has yet been validated in Mexico for adaptive behavior in adults. Sometimes American tests will be used after they have been translated into Spanish—for instance, the Vineland Scale for Children, or a non-standardized measure such as the Monterrey Test. Thus this case should be judged by a mixture of psychometric measures and qualitative information, with Mexican or American norms.

CONCLUSIONS

As recognized, the development of Mexican norms is an imperfect process, and many errors and mistakes could be identified in this effort. But do these errors invalidate the norms derived?

The validity evidence provided by the developer of the test in the United States cannot be challenged, since appropriate measures were taken to consider accurate translation, and linguistic variations were made of additional criteria to judge answers. Items were carefully revised and ordered according to the Facility Index in the standardization process, and directions about the administration of the test were clarified for the Mexican standardization.

Regarding norms, there is evidence of acceptable reliability, and there is a sense of a process that has amended parameters as a result of comments, revisions, and inconsistencies for use with different groups. The use of the 2004 revised norms is an example of the above. We agree that the analysis of current norms and the latest version of the test are absolutely necessary

to evaluate its advantages and limitations in testing the Mexican population, and as a result, we would expect new adjustments to the existing norms. Arguing that American norms are better to assess the Mexican population, without considering the complete processes of standardization and the latest revised norms, is both unfair and false.

The latest Mexican version of the test is more understandable to test takers, conveys clearer directions, and places items in a scale of difficulty that has proven to be empirically different to the American version. Most importantly, the Mexican WAIS-III provides norms elicited from the Mexican population; that is, it compares Mexicans against other Mexicans.

Suen and Greenspan (2008), in an effort to save a life, may have harmed unintentionally, the adaptation and validation of American tests in other countries. Adapting and validating tests in different cultural groups is an emerging and ill-financed practice that needs to be reassured and fostered.

Certainly, the latest version of the Mexican WAIS is imperfect; it was developed with limited funds and needs revisions and continuous monitoring, as do all standardized tests. Lessons learned in this thoroughly scrutinized version helped the test developers edit items, make adjustments to item order, and improved the translation. In addition, it helped us to realize many limitations to our data collection process. For instance, the experiences with the WAIS project in Mexico provided a template to avoid similar mistakes and unsuccessful practices during the standardization of similar tests, such as the standardization of the Wechsler Intelligence Scale for Children-4th edition (WISC-IV) carried out by the same investigators in 2006.

Finally, awareness should be given to the fact that, as a rule of thumb, it is unwise to make decisions and a final judgment about an individual, based upon the results of a single test. In the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement and Education, 1999) test users are “well advised, and may be legally required to consider other relevant sources of information on test takers, not just test scores” (p. 112). Thus, a capital punishment case decided on the interpretation of scores in a single test is not itself best practice.

Given these considerations, when the reader is challenged to assess a Mexican adult, choices are limited to: (a) the original American WAIS-III developed and translated to the Spanish with American norms, (b) the preliminary version of the Mexican WAIS-III, developed with norms related to a reference group of Mexican adults between 20 and 35 years old (Sanchez, 2001), or (c) the latest version of the Mexican WAIS-III with revised 2004 norms (Sanchez, 2003). The final decision is upon the user, who must indeed consider a number of contextual, developmental, and legal factors beyond the properties of the test itself.

REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement and Education (1999). *Standards for educational and psychological tests and manuals*. Washington, DC: American Psychological Association.
- Sanchez, P., Canton, B., & Sevilla, D. (1999). *Compendio de educación especial [Overview of special education]*. Cuauhtémoc, Mexico: Manual Moderno.
- Sanchez, P. (2001). Capítulo 6: Validación preliminar para México [Chapter 6: Preliminary validation for Mexico]. In D. Tulsy & J. Zhu (Eds.), *Manjua técnica del WAIS-III [Technical manual for the WAIS-III]* (pp. 123–138). Cuauhtémoc, Mexico: Manual Moderno.
- Sanchez, P. (2003). Capítulo 6: Validación preliminar estandarización a nivel nacional para México [Chapter 6: Preliminary validation and standardization at a national level for Mexico]. In D. Tulsy & J. Zhu (Eds.), *Manjua técnica del WAIS-III [Technical manual for the WAIS-III]*. Cuauhtémoc, Mexico: Manual Moderno.
- Sanchez, P., Canton, B., & Sevilla, D. (1999). *Compendio de Educación Especial*. Cuauhtémoc, Mexico: Manual Moderno.
- Suen, H., & Greenspan, S. (2008). Linguistic sensitivity does not require one to use grossly deficient norms: Why U.S. norms should be used with the Mexican WAIS-III in capital cases. *Psychology in Intellectual and Developmental Disabilities 12–15*. Official publication of Division 33, American Psychological Association. Available online at <http://www.apa.org/divisions/div33/docs%5C34-1.pdf>
- The Psychological Corporation (1997). *WAIS-III technical manual*. San Antonio, TX: Harcourt Brace and Company.
- Tulsy, D., & Zhu, J. (2001). *Escala Wechsler de Inteligencia para adultos-III*. Cuauhtémoc, Mexico: Manual Moderno.
- Tulsy, D., & Zhu, J. (2003). *Escala Wechsler de Inteligencia para adultos-III*. Cuauhtémoc, Mexico: Manual Moderno.
- Weiss, L. (2003). United States. In J. Georgas, L. Weiss, F. van de Vijver, & D. Saklofske (Eds.), *Culture and children's intelligence: Cross-cultural analysis of the WISC-III* (p. 50). San Diego, CA: Academic Press.