

# Reply to Sanchez-Escobedo and Hollingworth: Why the Mexican Norms for the WAIS-III Continue to be Inadequate

Hoi K. Suen

*Educational Psychology Program, Pennsylvania State University,  
University Park, Pennsylvania*

Stephen Greenspan

*Department of Psychiatry, University of Colorado, Denver, Colorado*

The discussion in Drs. Sanchez-Escobedo and Hollingworth's paper independently confirms virtually all our observations regarding the psychometric and interpretive deficiencies of the Mexican norms for very high-stakes decisions, such as that involved in an *Atkins* hearing. Test publishers have an ethical obligation to caution potential users against the premature use of a developing assessment that does not yet meet the needed precision and evidence of validity required for very high-stakes decisions.

In their article, Drs. Sanchez-Escobedo and Hollingworth (2009) provide some useful clarifications. What they present independently confirms virtually all our own (Suen & Greenspan, 2009) observations of the various deficiencies of the Mexican norms. The detailed subscale alpha values by age group presented in Table 1 of their paper does not explain the large standard error of measurement around the overall IQ scale score and its implicit low reliability value in practice—a contradiction that we have pointed out but which they did not address in their paper. The information in Tables 2 and 3 of their paper confirms that a disproportional rather than a proportional stratified sample was used. This is further confirmed by the use of the term “balanced sample” by these authors. The two tables describe characteristics of the sample used, but not the population. The implicit suggestion is for the user of the Mexican norms to induce the population from the sample. But what is the meaning of a score that is compared against an induced population with these particular characteristics, when such an induced population is

almost certainly a systematically biased subset of the unknown overall Mexican adult population?

We are delighted to learn that the one error in the scale score equation in the technical manual proved to be an inconsequential typographical error. However, all other errors are confirmed by Sanchez-Escobedo and Hollingworth. For example, they explain that the *df* value used for the  $\chi^2$  evaluation was 2. Yet, the correct *df* value for a  $2 \times 2$  table, which is the most probable case here, is 1. They also attempt to clarify that the interpretation error related to the  $\chi^2$  was an inadvertent but inconsequential misreporting of  $p \geq .001$ , and that the correct one should have been  $p \leq .001$ . This unfortunately compounds the problem since a  $\chi^2$  value of 4.86 in fact corresponds, for either 1 or 2 *df*, to a probability that is greater than .001.

Our observation that a linear standard scoring method rather than a normalized method was used by the test developer has been confirmed by Sanchez-Escobedo and Hollingworth as well. The resulting scores are thus not convertible to percentiles. Finally, and most importantly, in their discussion under *Lack of Representation of Certain Groups*, these authors clearly state: “...individuals with obvious mental retardation, for instance, may be better assessed using the specialized

---

Address correspondence to Hoi K. Suen, Pennsylvania State University, 103 CEDAR Building, University Park, PA 16802. E-mail: hoisuen@psu.edu

American norms rather than the Mexican Norms.” Therefore, the information provided by Sanchez-Escobedo and Hollingworth is supportive of our caution about the deficiencies of the Mexican norms and our recommendation against the use of such norms for the determination of mental retardation in capital cases.

Sanchez-Escobedo and Hollingworth cite a single publication about special education in general in Mexico as evidence that there is a definition of mental retardation in Mexico. While this represents the view of a team of three authors, it is hardly a professional consensual, legal, or official standard suitable for court decisions; nor can that possibly be in reference to the Mexican norms in question since the cited publication predates the establishment of the Mexican norms by six years.

The authors correctly point out that mental retardation is defined today in the US by a combination of three independent criteria from three different sources. Such definition is also adopted by the court. It is not correct, however, to suggest that having three independent sources should exempt us from requiring that each source of information should by itself be independently accurate, meaningful, and valid.

Of the three sources of information used to determine mental retardation, IQ scores bear an extra burden of proof of numerical precision and interpretability. IQ scores, being numbers, have an explicit claim and an implicit aura of numerical precision, along with a relatively narrow band of numerical cutoff definition. The ability to compare an IQ score against a numerical cutoff also provides an aura of impartiality. This explicit and implicit impression of precision, exactitude and impartiality gives IQ scores, rightfully or not, extra

weight when the three sources of evidence are evaluated. However, along with this extra weight, IQ scores have the additional obligation to demonstrate that these numerical values are accurate, meaningful, and are in fact interpretable and valid.

Ill-financed or otherwise, all assessment tools, systems and procedures necessarily go through successive stages of development and improvements. At any given stage, an assessment tool is appropriate for use for some purposes, but not for other purposes. While we appreciate the difficulty of test development within serious financial limits, test publishers have an ethical obligation to caution potential users against the premature use of an assessment still under development that does not yet meet the needed precision and evidence of validity required for very high-stakes decisions.

It is not true that the choice for assessing a Mexican adult is one of between good information from the US norms that have limited cultural sensitivity on the one hand or bad and uninterpretable information from the Mexican norms that are based on some biased sample of Mexican adults on the other. Bad and/or meaningless information under a mistaken cloak of accuracy and precision can be so misleading that it is better to have no information at all.

## REFERENCES

- Sanchez-Escobedo, P., & Hollingworth, L. (2009). Annotations on the use of the Mexican Norms for the WAIS-III. *Applied Neuropsychology*, 16(3), 223–227.
- Suen, H., & Greenspan, S. (2009). Serious problems with the Mexican norms for the WAIS-III when assessing mental retardation in capital cases. *Applied Neuropsychology*, 16(3), 214–222.