PRACTICAL RESOURCES for the Mental Health PROFESSIONAL

Culture and Children's Intelligence

Cross-Cultural Analysis of the WISC-III

Edited by
James Georgas
Lawrence G. Weiss
Fons J.R. van de Vijver
Donald H. Saklofske

AP

PRACTICAL RESOURCES for the Mental Health PROFESSIONAL

Psychologists and special educators are increasingly called upon to assess students newly immigrated from another country. The Wechsler tests are perhaps the most widely translated intelligence tests in the world and yet, little is known about the standardization efforts in different countries or how well the WISC-III travels across country, cultural and linguistic borders. This book informs professionals about these issues with respect to 16 different countries in which the WISC has been translated and validated for use.

Sources for obtaining translated versions are provided so that psychologists can assess immigrant students with greater confidence in multiple languages, and the assistance of a bilingual examiner. Issues presented are history of the development of the Wechsler tests, use of the WISC-III in each country and its potential use with ethnic groups in multicultural societies, and intelligence and cognitive processes from cross-cultural and indigenous perspectives. Relationships between WISC-III scores and affluence and education are also discussed.

The cross-cultural analysis of the data strongly indicates that the WISC-III is a remarkably robust measure of intelligence with cross-cultural relevance. It would appear that over fifty years of experience with the Wechsler tests and the periodic revisions during this period have resulted in a refined and valid measure of cognitive processes that has considerable power for assessing children's intelligence, even in different cultural contexts.

Culture and Children's Intelligence

Cross-Cultural Analysis of the WISC-III

James Georgas
The University of Athens, Greece

Fons J.R. van de Vijver
Tilburg University, The Netherlands

Lawrence G. Weiss
The Psychological Corporation, U.S.A.

Donald H. Saklofske
University of Saskatchewan, Canada

ISBN 0-12-280055-9

90038

9 780122 800559

Shelving Classification: Psychology

16

TAIWAN

HSIN-YI CHEN, YUNG-HWA CHEN, JIANJUN ZHU

PART III

CROSS-CULTURAL ANALYSIS OF THE WISC-III

17

PRINCIPLES OF ADAPTATION OF INTELLIGENCE TESTS TO OTHER CULTURES

FONS J. R. VAN DE VIJVER

18

METHODOLOGY OF COMBINING THE WISC-III DATA SETS

FONS J. R. VAN DE VIJVER, KOSTAS MYLONAS, VASSILIS PAVLOPOULOS, JAMES GEORGAS

18

METHODOLOGY OF COMBINING THE WISC-III DATA SETS

FONS J. R. VAN DE VIJVER
Department of Psychology
Tilburg University
Tilburg, The Netherlands

KOSTAS MYLONAS
Department of Psychology
The University of Athens
Athens, Greece

VASSILIS PAVLOPOULOS
Department of Psychology
The University of Athens
Athens, Greece

JAMES GEORGAS
Department of Psychology
The University of Athens
Athens, Greece

This chapter consists of three parts. In the first part an overview is given of which items were adapted and which items were closely translated per subtest in each of the countries. This presentation provides the background for the statistical analyses reported in Chapter 19; however, the overview is also interesting in its own right. It provides insight in the judgmental bias of the subtest, which refers to nonstatistical procedures to identify bias, based on a content analysis of the items. All local test development teams had to address two questions: (1) which American items were expected to be transferable to a new linguistic and cultural context without major alterations and (2) which items were assumed to require adaptations. As a consequence, country comparisons of the number of adapted items of the 11 subtests provide information about the judgmental bias in these subtests. The second part of this chapter describes (in a largely nontechnical way) the statistical analyses that are reported in Chapter 19. Conclusions are drawn in the third part.

OVERVIEW OF TEST ADAPTATIONS PER COUNTRY

In Chapter 17, a distinction was made between three ways of translating Wechsler Intelligence Scale for Children—Third Edition (WISC-III) subtests: applications (i.e., close translations of items), adaptations (i.e., change of item contents in order to enhance the suitability of an item for a particular cultural context), and assemblies (i.e., the development of a completely new instrument, needed when the original instrument would be entirely inappropriate in the new culture). Aggregated across countries, the vast majority of the items in the WISC-III adaptations described in this book, about 90%, has been closely translated or simply copied (in the case of pictorial stimuli), while a small minority of the items, about 10%, has been adapted. Assemblies were not used.

A detailed overview of the similarity of each subtest in each country to the U.S. subtest is presented in the Appendix. The tables indicate for each subtest whether the item was closely translated or adapted. In the case of a close translation, an item may appear in a different place in the item order. The order of the items is always determined by the empirically observed difficulty order. As a consequence, the order in a specific country may differ somewhat from the order in the U.S. subtest. The tables in the Appendix contain information about both the nature of the translation (application or adaptation) and the rank order of the closely translated items. So, the tables have three types of items: closely translated items with the same position in the item order in the U.S. as in a target country, closely translated items that have moved to a different place in the item rank order, and adapted items.

From a bias perspective, an interesting feature of the Appendix involves the proportion of adapted items across subtests and countries. An overview of these proportions is given in Table 18.1. The rows of the table present the country names in ascending order of proportions of adapted items. Analogously, the columns present the subtests in increasing order of their proportion of adapted items. A comparison of countries shows that, as could be expected, the smallest number of adaptations were found in the three English-speaking countries (Australia, Canada, and the UK). The largest proportions are found in Japan, The Netherlands and Flanders, and France. The rank order of the countries has some face validity in that culturally and geographically proximate countries tend to be close to one another. However, there was one notable exception. Whereas Korea and Taiwan are close to each other (with relatively low numbers of adapted items), Japan had the highest proportion of all countries. The reason behind the deviant position of Japan is not clear.

A column-wise comparison of Table 18.1 shows that the items of Object Assembly and Digit Span are identical in all countries. Block Design, Mazes, Picture Arrangement, and Picture Completion are identical in all countries except Japan; the test constructors in each country apparently judged that the performance subtests were culturally appropriate in their cultures. Thus, the Performance

stage the factorial agreement for all pairs of countries is computed. Four matrices, indicating the factorial agreement for each pair of countries, form the output of this stage. Finally, per factor a cluster analysis is employed to identify sets of countries with a high factorial agreement.

METRIC EQUIVALENCE

Assuming an affirmative answer to the question of the structural equivalence of the WISC-III, possibly amended for the fourth factor, we can continue with an analysis of metric equivalence. The adapted items may seem to preclude a direct comparison of raw scores across countries. The first and easiest solution would be to disregard the adapted items and to restrict the comparison to the items, used in all countries. A quick scan of the tables in the Appendix quickly reveals the unattractiveness of this option. If we would be forced to restrict the comparison to the common items, the comparisons would be based on very small item numbers, which would challenge the validity and replicability of the comparison. So, if we were to restrict the comparison to common items, some subtest comparisons not involving the U.S. may be based on small item sets.

The problem can be tackled using Item Response Theory (for introductions, see Fischer & Molenaar, 1995; Hambleton & Swaminathan, 1985; Hambleton, Swaminathan, & Rogers, 1991). Item Response Theory assumes that a person has a certain ability, which is statistically estimated on the basis of his or her responses on the subtest items; analogously, each item has a certain difficulty level, which is estimated on the basis on the number of correct responses to the item. The theory assumes that all items of a subtest measure the same underlying trait in all countries involved; in the case of cross-cultural data we also need to assume that items have the same item parameters in all countries involved (items should not be biased in favor of or against any country). Furthermore, Item Response Theory assumes local independence, which formally means that within groups of equal ability responses are statistically independent. In more informal terms, the assumption means that each subject answers each item independently and that there are no carry-over effects across items (e.g., effects due to memory or fatigue). If these assumptions are met, Item Response Theory allows for the estimation of a person's ability even if not all items are identical in all groups. As long as there are "anchors" (a set of items that are common to the countries to be compared), Item Response Theory can estimate the ability level of each person (and, by implication, the mean ability level of the country).

What does the application of Item Response Theory amount to in the present case? Let us take Vocabulary as an example. Suppose that two countries have 10 common and 20 country-specific items. We need to assume that in the two countries each of the 30 items measures the same underlying trait, say vocabulary knowledge (and nothing else than this trait). We also assume that the difficulty level of the common items is identical. The occurrence of an item that is much