

CRITICALLY EXAMINING LANGUAGE BIAS IN THE SOUTH AFRICAN ADAPTATION OF THE WAIS-III

CHERYL D FOXCROFT

SUSAN ASTON

cheryl.foxcroft@nmmu.ac.za

*Higher Education Access & Development Services
Nelson Mandela Metropolitan University*

ABSTRACT

In response to the growing demand for a test of cognitive ability for South African adults, the Human Sciences Research Council (HSRC) adapted the Wechsler Adult Intelligence Scales, third edition (WAIS-III) for English-speaking South Africans. The standardisation sample included both first and second language English speakers who were either educated largely in English or Afrikaans. The purpose of this article is to critically examine the adaptation process undertaken by the HSRC when standardising the WAIS-III for English-speaking South Africans by deliberating whether sufficient attention was paid to establishing if the measure was equivalent for various groups of English first and second language test-takers. In performing this critical examination, international test adaptation guidelines and standards, psychometric conventions, and national and international research findings were contemplated. The general conclusion reached was that the equivalence of the WAIS-III across diverse language groups has not been unequivocally established and there are indications that some bias may exist for English second language test-takers, especially if they are black or Afrikaans-speaking. Based on these conclusions, recommendations are made regarding the way forward.

Key words

Wechsler Adult Intelligence Scale, WAIS-III, language, test adaptation, test bias

The Wechsler Scales

Published in 1939, the Wechsler-Bellevue Adult Intelligence Scale was developed and standardised by David Wechsler as an alternative to the Stanford-Binet and with a clear purpose of measuring both verbal and non-verbal intellectual ability at the same time. The first major revision of the Wechsler-Bellevue was published in 1955 as the Wechsler Adult Intelligence Scale (WAIS). Further revisions resulted in the publication of the WAIS-R in 1988 and the Wechsler Adult Intelligence Scales, Third edition (WAIS-III) in 1997. During the latter and most recent revision, the test materials, item content and administration procedures were updated and three new subtests (Matrix Reasoning, Letter-Number Sequencing and Symbol Search) were added to the 11 subtests retained from the WAIS-R (Nell, 1999; Wechsler, 1997). While Full Scale, Verbal and Performance IQ scores are still computed, the third edition of the test also provides for grouping the scores of the subtests into more precise domains of cognitive functioning, namely: Verbal Comprehension Index, Perceptual Organisation Index, Working Memory Index, and Processing Speed Index. The introduction of the option of index scores aligns the WAIS-III with advances in neuropsychology and cognitive psychology and reduces testing time, as only those subtests necessary to evaluate a particular domain can be administered (Nell, 1999; Wechsler, 1997).

Although dialogue continues about the atheoretical nature of all editions of the Wechsler Intelligence Scales, they continue to be the most widely accepted and administered intelligence scales internationally, and there is little indication that their popularity will diminish significantly in the foreseeable future (Sparrow & Davies, 2000).

The Wechsler Intelligence Scales in South Africa

In South Africa, attention was initially focused on the adaptation of the Wechsler-Bellevue for the needs of South Africans, even though the test materials were not available in this country and data had to be obtained from Wechsler's manual, verbal descriptions and unscaled drawings (Huysamen, 1996; Nell, 1994). The adaptation began in 1947 and was initially undertaken by the Bureau of Personnel Research which became the National Institute of Personnel Research (NIPR) in 1948. Normative sampling across English and Afrikaans

language groups began in the 1950s. When the first major revision of the Wechsler-Bellevue was published as the Wechsler Adult Intelligence Scale in the United States of America in 1955, the NIPR found itself halfway through an expensive adaptation and norming exercise of the original measure. Instead of abandoning its adaptation of the Wechsler-Bellevue and concentrating on standardising the newly published WAIS, the NIPR decided to continue with the standardisation of the Wechsler-Bellevue. The wisdom of this decision has been seriously questioned (Nell, 1994).

The NIPR finally completed the adaptation and standardisation of the outdated Wechsler-Bellevue Scales for South Africa in 1969 and the adapted test was published as being the South African Wechsler Adult Intelligence Scale (SAWAIS). Despite the fact that most of the items were nearly three decades old, the South African version was named after the extensively revised version, the WAIS (Nell, 1994; Pieters & Louw, 1987). Although the finished product was considerably divergent from the original Wechsler-Bellevue on which it was based, it was not the WAIS, and did not incorporate the comprehensive modifications made to the Wechsler-Bellevue in 1955. The unintentional misnaming of the measure by the NIPR led to a belief among practising psychologists that the South African adapted Wechsler-Bellevue was, in fact, the WAIS, and this resulted in a reduction in the pressure to rapidly undertake another revision.

In 1987, Pieters and Louw publicly criticised the South African version of the Wechsler-Bellevue and exhorted the scientific, research and education communities to urgently consider replacing or re-standardising the test. Their critique cast aspersions on the quality of professional instruction in the field of psychology at local universities, and on the efficacy of the Professional Board for Psychology and the then Test Commission of the Republic of South Africa, who were responsible for ensuring that standards in psychometric testing were maintained at international levels (Nell, 1994). In addition, Pieters and Louw (1987) drew attention to the obligation of both the test publisher (the NIPR, which had subsequently been incorporated into the Human Sciences Research Council or HSRC) to clearly communicate to potential SAWAIS users and buyers that the measure was based on the Wechsler-Bellevue and

not on the WAIS, and that the statistical properties of the SAWAIS were largely unknown. At that stage the norms of the SAWAIS were so outdated that the scores were no longer valid predictors of deficits or normality (Nell, 1994).

As the SAWAIS became increasingly dated, criticism of its continued use escalated (Claassen, Krynauw, Holtzhausen, & Mathe, 2001a). Questions were raised and extensive evidence was collected, the result being a clear indication that the ongoing administration of the South African version of the Wechsler-Bellevue was not in the best interests of the public (Nell, 1994). In some cases practising psychologists resorted to using the USA revision of the WAIS-R, which had no South African norms, or the revised Senior South African Individual Scale (SSAIS-R) which had only been normed for school-going children (Claassen et al., 2001a).

During the academic boycott period in the late 1980s and early 1990s, the Psychological Corporation rejected attempts by the HSRC to initiate standardisation of the WAIS-R for South Africans. However, once sanctions and boycotts were lifted after the demise of Apartheid and the formation of a democratic South Africa in 1994, the HSRC signed a contract with the Psychological Corporation in December 1997 to adapt and standardise the WAIS-III for English-speaking South Africans (Claassen, Krynauw & Holtzhausen, 2000).

The adaptation of the WAIS-III for English-speaking South Africans

The primary objective of the process of adapting the WAIS-III was to add or adapt items that were considered to be more relevant to the South African context and to develop norms for English-speaking South Africans from the four main cultural groups, namely, blacks, coloureds, Indians, and whites (Claassen et al., 2001a). In view of the trend towards the use of English by urbanised African-language speakers, the advisory committee suggested that each of the four cultural groups should constitute 25% of the standardisation sample. By allocating 25% to each group, investigation into differential item functioning (DIF) was facilitated (Claassen et al., 2001a). Furthermore, given that both first and second language English speakers were included in the standardisation sample, it was important to establish the English proficiency levels of the participants. Consequently, all participants were required to write a short English test which made it possible to compare the levels of English proficiency between the groups.

Both quantitative and qualitative methods were used to identify which items needed to be adapted or replaced. The original items were administered to a multicultural sample of black (N=165), coloured (N=230), Indian (N=191) and white (N=203) adults. DIF analyses were performed to guide the evaluation of each item. Items that were found to be biased against non-white South Africans were replaced.

From a qualitative perspective, items in the Information and Comprehension subtests that appeared at face value to be foreign to South African cultures were replaced with more relevant items (Claassen et al., 2001a). In addition, the opinions of 11 knowledgeable experts from all cultural groups were sought regarding the suitability of each item in the Vocabulary, Information and Comprehension tests for English-speaking South Africans (Claassen et al., 2001a). Details of the comments of the expert review group have not been reported in the technical report, nor are any indications provided as to how comments were incorporated into the adaptation of the items. This omission casts doubt on the significance accorded to qualitative data by the test adaptors. In addition, no indication is provided whether the qualitative and quantitative data were used in combination to determine whether an item should have been adapted. The lack of reporting by the project team on the qualitative data suggests that the process of standardising the WAIS-III relied heavily on quantitative information.

According to the task group assigned by the HSRC to adapt the measure to the South African context, very little item bias was detected, resulting in minimal changes to the original test. Modifications were made to the Vocabulary, Information, Arithmetic and Comprehension subtests. Readers are referred to Claassen et al., (2001a) and Claassen, Krynauw, Paterson and Mathe (2001b) for a detailed discussion of the specific changes made.

Unfortunately, after adapting certain items that proved to be biased, the new items were not re-piloted in order to determine the possible existence of continuing cultural or language bias. This requirement is stipulated in the guidelines for test adaptation published by the International Test Commission (ITC) (International Test Adaptation Guidelines [On-line], 2000). Guideline D.4, Test Development and Adaptation states that "Test developers/publishers should provide evidence that item content and stimulus materials are familiar to all intended populations."

Although the measure was standardised ostensibly only for English-speaking South Africans, the test adaptation team needed to be conscious of the fact that culturally diverse backgrounds as well as differing levels of English proficiency among test-takers would have a marked impact on the understanding of and familiarity with English terms and on test performance. As will be argued in the next section, there is some doubt whether sufficient attention was paid to this matter during the adaptation process and whether it can be confidently asserted that the South African adaptation of the WAIS-III is not biased against second language English speakers.

Impact of language on test performance

Language as a mediator of test performance

Language is one of the parameters along which cultures vary. In South Africa, 11 official languages are recognised: nine African languages, Afrikaans and English. English-speaking learners are educated through the medium of English, while in most cases African language learners are educated in their home language until they reach Grade 4 and thereafter mainly in English. Most Afrikaans-speaking learners are educated in Afrikaans, and study English as a subject at school. Consequently, there is a widely held view in South Africa that as the language of learning from Grade 4 onwards is either English or Afrikaans, test performance will not be negatively affected if test-takers are assessed in their language of teaching and learning (Claassen et al., 2001b). Furthermore, the dominant language in business and industry is English. Consequently, when administering an individual intelligence scale like the WAIS-III, psychologists often argue that it is justifiable to administer the measure in English, irrespective of whether English is the first or second language of the test-taker, as it is important that test-takers can demonstrate their ability to perform test tasks in the language that will be used in the workplace (Koch, 2005). However, Koch (2005) is critical of this approach for two reasons. First, it assumes that the scores on the measure are comparable across language groups. Second, it ignores the fact that language can be a 'nuisance factor' that impacts on the test performance of English second language speakers.

Language may be the most important mediator of test performance, especially when the language in which the measure is administered is not the home language of the test-taker. Concepts may be denied or alternatively, made available, to test-takers who are native or non-native speakers of the language of administration (Nell, 1994). The use of colloquial or archaic language in test items can lead to misunderstanding and miscommunication by test-takers, which may ultimately influence scores negatively (Nell, 1999). Herbst and Huysamen (2000) found that environmentally disadvantaged children assessed in a language other than that spoken at home performed at a significantly lower level than those who were assessed in their mother tongue. Items involving verbal comprehension

were found to be biased against test-takers who spoke an African language at home, even though they had been exposed to English on a daily basis.

Language usage and reading ability can significantly impact on test scores when measures are administered in languages or with cultures other than those for which the test has been standardised. Individuals who do not read test items accurately and those who fail to understand the content of a test item are more likely to respond incorrectly (Hinkle, 1994; Shuttleworth-Edwards, Kemp, Rust, Muirhead, Hartman & Radloff, 2004). Additionally, it has been found that language is one of the primary influencers of intelligence test performance, and can have a significant negative impact when a test is administered in the test-taker's second or third language (Nell, 1999).

Language and performance on the South African adaptation of the WAIS-III

The task team responsible for the adaptation of the WAIS-III for English-speaking South Africans stated that "it is unlikely that performance in some of the performance subtests will be adversely affected by the language used by the tester, even if the subject has only a limited command of that language, providing he/she has clarity on what is expected of him/her" (Claassen et al., 2001a, p. 8). This comment is contrary to the views and findings of Koch (2005), Herbst and Huysamen (2000), Hinkle (1994), Nell (1994, 1999) and Shuttleworth-Edwards et al., (2004) expressed above. While test-takers whose first language is not English may understand the wording of items, the interpretation of meaning varies significantly across cultures and first and second language English speakers, and may well impact negatively on test scores.

Aston (2006) obtained the views of psychology professionals in the Eastern and Western Cape on each of the WAIS-III subtests regarding whether items were potentially problematic (biased) for English, Afrikaans and Xhosa speakers. Her results indicated that problems were experienced by test-takers with the language of certain of the verbal and performance subtests. When it came to the performance subtests, the participants reported that Xhosa- and Afrikaans-speaking test-takers were confused by the wording used in the instructions of the Picture Completion, Digit Symbol-Coding, and Block Design subtests. The test users themselves experienced difficulty with the instructions of the Matrix Reasoning subtest and recommended that these be simplified as well as translated into Afrikaans and Xhosa in order to facilitate administration of the subtest. Aston's (2006) findings suggest that the test adaptation team should have spent more time qualitatively and quantitatively evaluating the performance subtests and their instructions and possibly introducing some adaptations. Hambleton and De Jong (2003) concur with this view in that they argue that producing a test which is linguistically appropriate for use in more than one language involves adaptation of both verbal and non-verbal tasks.

What information did the test adaptation team provide regarding whether second language English speakers are disadvantaged by taking the measure in English? To their credit, the team performed various investigations and analyses to explore the factor structure and compare the performance of various cultural and language groups.

Factor structures were derived using principle factor analysis with varimax and oblique rotation for blacks, coloureds, Indians and whites. In addition, factor structures were derived for blacks with and African language as a mother tongue, Afrikaans-speakers who work in an English environment, and for Afrikaans-speakers who speak Afrikaans at work. Claassen et al. (2001a and b) reported that the solutions derived supported the four index scores for English-speaking coloureds, Indians and whites. However, the solutions derived for black English speakers and African language speakers as well as for the two

Afrikaans-speaking groups were more mixed and less satisfactory. In particular, Claassen et al., (2001a and b) concluded that there was weak support for the structure of the index scores for the black samples. The findings of the factor analysis raise questions regarding the equivalence of the WAIS-III across cultural and language groups and for test-takers whose home language is not English but who take the test in English. It is a pity that Claassen et al. (2001a and b) did not statistically compare the similarity of the factor structures for the various groups by, for example, computing coefficients of congruence. This would have provided empirical information regarding the equivalence (similarity) of the factor structures and could have indicated whether there was greater similarity for some of the index scores than for others. With the wisdom of hindsight, it should also be questioned whether merely deriving factor solutions for various samples was sufficient to reach a conclusion regarding the equivalence of the WAIS-III across various language and cultural groups. It is increasingly being argued in the literature that multiple methods should be used to evaluate structural equivalence as opposed to only using one method and that matched groups should ideally be used (Koch, 2005; Sireci & Khaliq, 2002). Researchers suggest that a combination of methods such as principal components analysis, weighted multidimensional scaling, and structural equation modelling, together with various methods for exploring differential item functioning (DIF), should be used to comprehensively evaluate the structural equivalence of a measure across language versions or language groups (Koch, 2005; Sireci & Khaliq, 2002). Only one method (principle factor analysis) was used to investigate the equivalence of the WAIS-III for various cultural and language groups and matched samples were not used. Users of the WAIS-III should thus be aware that there is not yet sufficient and unequivocal evidence to conclude that the measure is structurally equivalent across cultural and language groups.

Other than investigating the factor structure of WAIS-III across various groups, the performance of different groups formed on the basis of language and culture was also explored. It should be noted that Claassen et al., (2001a and b) only reported means and standard deviations for different groups and did not provide any inferential statistics, which would have aided the interpretation of whether differences among the various groups were statistically significant or not. Consequently, the discussion on the performance of the various groups that will be highlighted here remains speculative, although the present authors used the means, standard deviations and sample sizes provided by Claassen et al., (2001b) to test whether the means differed significantly. When testing whether two means differ significantly using the STATISTICA package, only p values are provided. Hence, reference will only be made to p values when commenting on whether two means differed significantly or not.

The following mean scores were obtained on the English proficiency test: 112.62 (SD = 9.75) for English-speaking whites, 106.52 (SD = 11.19) for English-speaking coloureds, 102.97 (SD = 1.50) for Afrikaans-speaking whites, 99.72 (SD = 13.56) for English-speaking blacks, 95.88 (SD = 14.75) for Afrikaans-speaking coloureds, and 90.37 (SD = 17.47) for blacks who spoke English at work but for whom English was largely a second or third language. When the mean differences were statistically compared it was found that the mean score for English-speaking whites was significantly higher than that of Afrikaans-speaking whites ($p < .001$), English-speaking coloureds ($p = .0012$), Afrikaans-speaking coloureds ($p < .001$), English-speaking blacks ($p < .001$) and blacks who spoke English at work although this was not their first language ($p < .001$). The groups thus differed significantly in terms of their English proficiency levels, with groups for whom English was a second (or third) language scoring 0.5 to almost two standard deviations below the white English first language sample.

Table 1 provides information on the index scores obtained by various language and cultural groups. This table was compiled from statistical information in Tables 9.8 and 9.10 in Claassen et al., 2001b (p. 63 and 66). Furthermore, Table 2 contains the p-values obtained when testing whether the difference between the means for the white English-speaking sample and those of the other language and cultural samples was significantly different.

TABLE 1
MEANS AND STANDARD DEVIATIONS FOR INDEX SCORES FOR
VARIOUS LANGUAGE AND CULTURAL GROUPS

	N	Verbal Compre- hension	Perceptual Organi- sation	Working Memory	Processing Speed
English-speaking whites	70	109,40 (15,49)	111,51 (15,42)	104,96 (14,73)	107,13 (15,90)
Afrikaans-speaking whites	97	102,88 (14,10)	110,72 (16,37)	101,07 (12,87)	108,32 (15,15)
English-speaking coloureds	72	103,93 (13,70)	102,04 (11,65)	98,67 (14,33)	101,64 (13,87)
Afrikaans-speaking coloureds	96	92,36 (13,44)	93,89 (14,49)	94,22 (13,03)	97,17 (13,74)
English-speaking blacks	35	101,71 (15,02)	97,63 (16,84)	96,23 (11,90)	96,97 (13,99)
Blacks who speak English at work	196	91,47 (14,69)	88,17 (14,49)	91,42 (14,10)	88,60 (9,80)

(SD provided in parentheses)

TABLE 2
SIGNIFICANCE LEVEL (P) WHEN COMPARING THE MEANS
OF THE WHITE ENGLISH-SPEAKING SAMPLE TO THE MEANS
OF THE OTHER LANGUAGE AND CULTURAL GROUPS

English-speaking whites compared with:	Verbal Compre- hension	Perceptual Organi- sation	Working Memory	Processing Speed
Afrikaans-speaking whites	0,001	0,753	0,072	0,624
English-speaking coloureds	0,027	<0,001	0,011	0,030
Afrikaans-speaking coloureds	<0,001	<0,001	<0,001	<0,001
English-speaking blacks	0,017	<0,001	0,003	0,002
Blacks who speak English at work	<0,001	<0,001	<0,001	<0,001

When comparing the index scores of the differing combinations of language and cultural groups provided in Table 1, it is clear that the English-speaking white sample consistently obtained higher mean scores than the other groups. From Table 2 it can be seen that the mean for the white English-speaking group was significantly ($p < .05$) higher than that of all the other groups for Verbal Comprehension. In addition, the mean Perceptual Organisation, Working Memory and Processing Speed index scores of the white English-speaking sample was significantly higher than the means for all the groups, except for the Afrikaans-speaking white group. The differences in means between white English speakers and the other groups ranged between 6 and 17 points and in the majority of cases the mean differences were statistically significant. This raises questions about the potential differential impact of English proficiency on WAIS-III performance.

Some may argue that the comparisons of the test performance of the different language and cultural group combinations could

be attributable more to differences between cultural groups than to language proficiency. While there might be some validity to this argument, the comparison of the mean performance of English and Afrikaans speaking whites and coloureds provided in Table 1 suggests otherwise. Not only did English-speaking coloureds consistently obtain significantly higher mean scores than Afrikaans-speaking coloureds (p-values ranged between .038 and <.001), but in the case of the Verbal Comprehension and Perceptual Organisation scores the mean difference was close to 10 points, which is a considerable difference that was found to be statistically significant ($p < .001$) in both instances. Although English-speaking whites generally obtained higher mean scores than Afrikaans-speaking whites, the difference was small. However, the Verbal Comprehension mean for second language English speakers was almost 7 points lower than that of first language English speakers and this difference was found to be statistically significant ($p = .005$). Thus, on the index score where language plays a significant role in terms of the nature of the test tasks and the constructs being tapped, it is worrying that the scores of second language English speakers were significantly depressed. Comparison of the performance of different language groups within the same cultural group thus leads one to the conclusion that the impact of language on test performance cannot be ignored and cannot solely be explained in terms of cultural differences in cognitive test performance.

As education has been highlighted as being a critical moderator of test performance (Claassen, et al., 2001a and b; Shuttleworth-Edwards, et al., 2004), another argument that could be put forward is that the differences between the various cultural and language groups could best be explained in terms of differences in the levels of education and the quality of the education received rather than in terms of differing levels of English proficiency. Claassen et al. (2001b) did not report on analyses in which the main and interaction effects of educational level, language and culture on WAIS-III performance were explored. However, when Shuttleworth-Edwards, et al. (2004) explored the impact of quality and level of education on WAIS-III test performance, they also incidentally provided information on the performance of educationally comparable samples of English and African first language speakers. English first language graduates respectively obtained mean Verbal, Performance and Full Scale IQ scores of 124.93, 116.14 and 123, while African first language graduates obtained mean scores of 116.10, 107.80, and 113.40 respectively. While the results of the groups were not statistically compared by Shuttleworth-Edwards, et al. (2004), the present authors tested the mean differences for significance. While there was no significant difference between the Performance IQ scores for the two groups ($p = .07$), the Verbal and Full Scale IQ scores of the English first language graduates were significantly higher than those of the African first language graduates ($p = .01$ in each instance). The significantly depressed scores of African first language speakers who were tested in English is a source for concern. Furthermore, these results once more suggest that the impact of language on test performance is a real issue that has to be specifically addressed.

One way in which the differential impact of language on test performance could have been addressed would have been to explore whether separate norms should have been developed for different language and cultural group combinations. Claassen et al. (2001b) contemplated this possibility but decided to weight the cultural groups in terms of quality of education and not to attempt to also develop norms for the different language subgroups. They argued that if the language of learning was taken into account, this could justify testing blacks who have an African home language in English. However, as Afrikaans speakers educated in Afrikaans performed poorly on the English version of the WAIS-III, Claassen et al. (2001b) suggested that the measure should be

administered in Afrikaans to this group of test-takers. To this end they developed an Afrikaans translation of the verbal subtests but cautioned that the “translated version has as yet not been standardised and should be used with this knowledge in mind” (p. 73). The ITC’s test adaptation guidelines (International Test Adaptation Guidelines [On-line], 2000) as well as the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 1999) clearly indicate that when a measure is translated and/or adapted, evidence needs to be provided regarding the equivalence of the translated/adapted versions. It is thus unacceptable that Claassen et al., (2001b) provided an Afrikaans translation of the verbal subtests without also providing information regarding the equivalence of the translation. Until such information is available, psychologists who follow good assessment practices should refrain from using the Afrikaans version of the WAIS-III. To amplify this suggestion, the findings of two recent studies are pertinent.

Grieve (2005) identified certain problematic items in the Afrikaans translation and was critical of the fact that no scoring criteria were provided for the Afrikaans translation. These findings and sentiments were echoed by Aston (2006). For example, in Aston’s (2006) study, participants, who were all psychology professionals, commented qualitatively that on the Vocabulary Subtest: “The meaning of the items used in the Afrikaans translation differs from the meanings of the original English items. Afrikaans people generally perform less well on these items” (p. 96) and that “Because the meanings of these words differ significantly from the original English version and are less commonly used, Afrikaans test-takers are likely to perform less well on this subtest. The questions and the scoring criteria need to be translated into Afrikaans in order to obtain parity and equivalence” (p. 96). There are thus sufficient indications that there are problems with the Afrikaans translation of the WAIS-III which psychologists should take heed of. It should also be noted that no norms have been developed for Afrikaans speakers who are assessed using the Afrikaans version of the WAIS-III for the verbal subtests, which is also problematic.

DISCUSSION

When judged against international guidelines for test adaptation and for using measures across linguistically diverse groups, certain problems have been identified in the South African adaptation of the WAIS-III. The adapted version seems to be appropriate for white English-speaking South Africans. However, the results of the factor analyses as well as the comparisons of performance among the groups strongly hint that there are some doubts regarding the equivalence of the WAIS-III across various language groups and that the measure may be biased against English second language black test-takers and coloured and white Afrikaans-speaking test-takers.

The Way Forward

It would be inappropriate to merely be critical of the adaptation of the WAIS-III for first and second language English speakers without offering some suggestions regarding how language issues can be addressed in the WAIS-III. Consequently, in the concluding section of this article suggestions will be offered related to the responsibility of the test distributor, the psychologists using the WAIS-III, researchers, and the Psychometrics Committee of the Professional Board for Psychology with respect to addressing this matter.

Responsibility of the test distributor

The HSRC’s WAIS-III test adaptation team was disbanded when the project ended and the onus now rests on the test distributor

in South Africa to address limitations in the South African adaptation. The test distributor should specifically focus on:

1. Undertaking a further, more thorough, qualitative bias review of all the subtests of the WAIS-III, including the performance subtests to detect items and instructions that require adaptation or replacement. The review team should consist of psychologists, measurement experts, linguists and anthropologists. The team should furthermore be representative of English first and second language speakers from the four major cultural groups. The recommendations of Aston (2006) regarding items that might be biased against second language English speakers and problematic instructions could be used to facilitate the review process.
2. Establishing the equivalence of the measure across the diverse language groups that it is intended to be administered to. As suggested by Koch (2005) and Sireci and Khaliq (2002) a combination of statistical methods should be used when equivalence is investigated.
3. Re-examining whether separate norms for various combined language and cultural groups might result in the measure being able to be used more fairly with English second language test-takers in particular.
4. Refining the Afrikaans translation of the verbal subtests and using judgemental and empirical methods to establish the equivalence of the English and Afrikaans versions. In addition, norms should be developed for Afrikaans-speaking South Africans.
5. Exploring the possibility of adapting/translating the measure into various African languages. Once an equivalent Afrikaans translation is available, it will be difficult to justify why only an Afrikaans translation is provided. As Hambleton and De Jong (2003, p. 130) observe, “Growing recognition of multiculturalism has raised awareness of the need to provide for multiple language versions of tests and instruments intended for use within a single national context”. Having the WAIS-III available in multiple languages will allow psychologists to assess test-takers in the language in which they are most proficient.

Responsibility of psychologists using the WAIS-III

Psychologists who use the WAIS-III should critically study the technical report (Claassen et al., 2001b) to reach their own conclusions regarding the limitations of the adapted South African version. In addition, they should seek out reviews of the measure and South African research studies. This should provide them with sufficient information regarding test-takers for whom it is appropriate and inappropriate to administer the measure to. The acid test will be whether psychologists will follow good assessment practice guidelines and not use the WAIS-III in instances where there is insufficient information to support the possibility that the results will be valid and that language factors did not negatively impact on test performance.

Responsibility of researchers

It is essential that South African researchers critically examine the adaptation of the WAIS-III and explore the psychometric properties of the WAIS-III for diverse groups. From their findings, refinements to the measure and enhancements to the use and interpretation of WAIS-III test performance can be suggested. By way of example, three South African studies on the WAIS-III were reported in this article. Namely, studies by Aston (2006), Grieve (2005) and Shuttleworth-Edwards, et al., (2004). While there are also other South African studies that have researched the WAIS-III, more are needed so that a substantial body of knowledge can be developed.

Responsibility of the Psychometrics Committee

The Psychometrics Committee of the Professional Board for Psychology are tasked with advising the Board on matters pertaining to psychological tests and testing and to classify psychological tests. When a measure is submitted for classification, two independent reviewers evaluate it to

determine whether use of the measure constitutes a psychological act, whether application of the measure and/or its results could have harmful consequences, how appropriate the measure is for the multicultural, multilingual South African context, and whether the psychometric properties of the measure have been comprehensively evaluated. As the WAIS-III is on the list of tests classified as being a psychological test by the Psychometrics Committee, it must be assumed that the measure was evaluated for classification purposes. Nonetheless, this process failed to identify critical issues regarding the appropriateness of the WAIS-III for diverse language groups, the lack of sufficient information regarding the equivalence of the measure across various groups, the appropriateness of the norm groups, and the availability of an unsubstantiated Afrikaans version of the verbal subtests with no norms for Afrikaans speakers. One explanation for why this happened might be that the classification and review criteria need to be expanded and are not explicit and detailed enough. It is encouraging to note that the test classification methodology and criteria used are currently being contemplated by the Psychometrics Committee with the view to possible refinement. Psychologists depend on this committee to ensure that the psychological tests that they use comply with quality and psychometric standards. It is thus further imperative that the Psychometrics Committee subjects its refined test classification and review process and criteria to international benchmarking.

Concluding remark

The test adaptation process is fraught with difficulties and it is through the combined efforts of test users and researchers that limitations in the adapted version of a measure can be brought to the attention of test developers and distributors. It is thus hoped that this article has made a constructive contribution to the further refinement and adaptation of the WAIS-III as regards minimizing the impact of language on test performance.

REFERENCES

- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- Aston, S. (2006). *A qualitative bias review of the adaptation of the WAIS-III for English-speaking South Africans*. Unpublished Master's dissertation, Nelson Mandela Metropolitan University, Port Elizabeth.
- Claassen, N.C.W., Krynauw, A.H. & Holtzhausen, H. (2000). *Standardising the Wechsler Adult Intelligence Scale-Third edition (WAIS-III) for South Africa*. Pretoria: Human Sciences Research Council.
- Claassen, N.C.W., Krynauw, A.H., Holtzhausen, H., & Mathe, M. (2001a). *Wechsler Adult Intelligence Scale-Third edition: performance of South African reference groups*. Pretoria: Human Sciences Research Council.
- Claassen, N.C.W., Krynauw, A.H. & Paterson, H., & Mathe, M. (2001b). *A standardization of the WAIS-III for English-speaking South Africans*. Pretoria: Human Sciences Research Council.
- Grieve, K. (2005). *Use of the WAIS-III for Afrikaans-speaking South Africans*. Paper delivered at the 11th annual congress of the Psychological Society of South Africa, Cape Town, 20-23 September 2005.
- Hambleton, R.K. & De Jong, J.H.A.L. (2003). Advances in translating and adapting educational and psychological tests. *Language Testing*, 20 (2), 127-134.
- Herbst, I. & Huysamen, G.K. (2000). The construction and validation of a developmental scale for environmentally disadvantaged preschool children. *South African Journal of Psychology*, 30 (3), 19-24.
- Hinkle, J.S. (1994). Practitioners and cross cultural assessment: a practical guide to information and training. *Measurement and Evaluation in Counselling and Development*, 27 (2), 748-756.
- Huysamen, G.K. (1996). *Psychological measurement: an introduction with South African examples* (3rd ed.). Pretoria: J.L. Van Schaik Publishers.
- International Test Commission (2000). *International test adaptation guidelines* [On-line]. Available: <http://www.intestcom.org>. Accessed June, 2006.
- Koch, E. (2005). *Evaluating the equivalence, across language groups, of a reading comprehension test used for admissions purposes*. Unpublished doctoral thesis, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa.
- Nell, V. (1994). Interpretation and misinterpretation of the South African Wechsler-Bellevue Adult Intelligence Scale: a history and a prospectus. *South African Journal of Psychology*, 24 (2), 100-108.
- Nell, V. (1999). Standardising the WAIS-III and WMS-III for South Africa: legislative, psychometric, and policy issues. *South African Journal of Psychology*, 29, 128-137.
- Pieters, H.C. & Louw, D.A. (1987). Die Suid-Afrikaanse Wechsler-Intelligensieskaal vir volwassenes: 'n kritiese perspektief. *South African Journal of Psychology*, 17, 145-149.
- Shuttleworth-Edwards, A.B., Kemp, R.D., Rust, A.L., Muirhead, J.G.L., Hartman, N.P. & Radloff, S.E. (2004). Cross-cultural effects on IQ test performance: A review and preliminary normative indications on WAIS-III test performance. *Journal of Clinical and Experimental Neuropsychology*, 26 (7), 903-920.
- Sireci, S.G. & Khaliq, S.N. (2002). *Comparing the psychometric properties of monolingual and dual language test forms*. (Center for Educational Assessment Research No. 458). Amherst, MA: School of Education, University of Massachusetts, Amherst.
- Sparrow, S.S. & Davies, S.M. (2000). Recent advances in the assessment of intelligence and cognition. *Journal of Child Psychology & Psychiatry*, 41 (1), 117-131.
- Wechsler, D. (1997). *WAIS-III administration and scoring manual*. San Antonio, Texas: Psychological Corporation.

REVIEW PANEL

Dr Kate Cockroft	University of the Witwatersrand
Prof Marié de Beer	University of South Africa
Prof Deon de Bruin	University of Johannesburg
Dr Karina de Bruin	University of Johannesburg
Prof Robert Doktor	University of Hawaii at Manoa
Prof Cheryl Foxcroft	Nelson Mandela Metropolitan University
Dr Dirk Geldenhuys	University of South Africa
Prof Kate Grieve	University of South Africa
Prof Gert Huysamen	Gordon Institute of Business Science & University of the Free State
Dr Martin Jooste	University of Johannesburg
Prof Wilhelm Jordaan	University of Pretoria
Dr Charlene Lew	Gordon Institute of Business Science
Mr Deon Meiring	South African Police Services
Prof Ian Rothmann	North-West University
Dr Hilton Rudnick	Private Practice
Prof Pieter Schaap	University of Pretoria
Prof Faans Steyn	North-West University
Prof Callie Theron	Stellenbosch University
Prof Esmé van Rensburg	North-West University
Prof Delene Visser	University of Johannesburg