

Development and Reliability of the Indonesian Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV)

Christiany Suwartono^{1,2}, Magdalena S. Halim¹, Lidia L. Hidajat¹, Marc P. H. Hendriks^{2,3}, Roy P. C. Kessels^{2,4}

¹Department of Psychology, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia

²Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

³Department of Behavioural Sciences, Academic Centre of Epileptology, Kempenhaeghe, Heeze, The Netherlands

⁴Department of Medical Psychology, Radboud University Medical Center, Nijmegen, The Netherlands

Email: christiany.suwartono@atmajaya.ac.id

Received 8 July 2014; revised 5 August 2014; accepted 29 August 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Through the years, several translated versions of Wechsler's intelligence test have been used in Indonesia, in clinical, educational or industrial settings. However, instruments such as Wechsler-Bellevue Intelligence Scale are outdated, have not been validated and lack proper normative data, resulting in measurement errors and invalid decisions made on the intellectual potential of individuals. The primary aim of this study was to adapt and validate the Wechsler Adult Intelligence Scale—fourth edition (WAIS-IV) for use in Indonesia. We described the first phase in the adaptation of the WAIS-IV in the Indonesian language, including translation, item analysis, and reliability of the subtests. The sample of this research consisted of 148 healthy participants who are representative for the Indonesian population with respect to gender, age groups (ages 16 to 83), educational levels, and ethnic background. Results showed that the sequence of the US WAIS-IV cannot be applied in Indonesia due to differences in index difficulties. Cronbach's coefficient alphas for the WAIS-IV subtests ranged from .74 - .92. For the subtests from the Verbal Comprehension Index, the inter-rater agreement ranged between .91 - .97. In all, the adaptation of the WAIS-IV for Indonesia is psychometrically promising.

Keywords

Intelligence, Assessment, Item Analysis, Reliability

1. Introduction

Assessment of intelligence is one of the most important topics in psychological testing. Worldwide, Wechsler's intelligence tests are best known and the most widely used for testing intelligence. The original version of the Wechsler intelligence test, the Wechsler-Bellevue Intelligence Scale (WBIS) for adolescents and adults was developed in 1939. Internationally, the WBIS has been extended and modified several times to WAIS, WAIS-R, WAIS-III, and the recently published WAIS-IV (Wechsler, 2008a, 2008b). Adaptations from the original US versions of these tests have been published in many countries, such as the Netherlands (see Van der Heijden, Van den Bos, Mol, & Kessels, 2013), Japan (see Murayama, Iseki, Tagaya, Ota, Kanasuki, Fujishiro, Arai, & Sato, 2013), Finland (see Roivainen, 2013), France (see Lecerf, Golay, & Reverte, 2012), South Africa (see Grieve & Van Eeden, 2010), Canada (see Lange, 2007), Spain (see Melendez, 1994), and China (see Lynn & Dai, 1993). In the latest WAIS-IV items and subtests have been substantially revised, new subtests have been added and norms have been updated to take cohort effects into account (Wechsler, 2008b). The WAIS-IV updates also incorporated theoretical advances in neuropsychology, cognitive neuroscience, and contemporary intelligence theory, as well as increasing sophistication in psychological measurement (Weiss, Saklofske, Coalson, & Raiford, 2010).

In Indonesia, the WBIS is still being used for intelligence testing. However, details about its translation (presumably done in the 1960s or 1970s), standardization, psychometric properties and the development of the written manual are unknown, and this version is probably unauthorized (Lembaga Pengembangan Sarana Pengukuran dan Pendidikan Psikologi Fakultas Psikologi Universitas Indonesia, n.d.) and certainly outdated. Despite these shortcomings, the WBIS is still being taught and the most widely administered intelligence test in Indonesia. A translated version of the later Wechsler Adult Intelligence Scale is also available (Seksi Psikodiagnostik, 1992), but suffers from the same shortcomings, is also probably unauthorized and has never gained wide acceptance in the Indonesian psychology community. As an example, in Bandung, several faculties of psychology preferred the WBIS over the WAIS as the translated WAIS items were culturally biased (Polhaupessy, n.d). The lack of valid intelligence tests results in measurement error, misleading decisions made about the potentials of individuals (i.e., in vocational training or job selection), as well as clinical misdiagnosis. This stresses the urgent need for the development and adaptation of a valid intelligence test for the Indonesian population. The WAIS-IV can be used to identify intellectual disability, intellectual giftedness, and cognitive functioning in examinees (Wechsler, 2008a). The WAIS-IV also can be used as part of neuropsychological assessment in clinical settings (Fitrikasari, Jaeri, & Bintoro, 2013). Therefore, the focus of this study is the development of the Indonesian version of the WAIS-IV and to psychometrically evaluate the Indonesian WAIS-IV.

The WAIS-IV is an individually administered test battery for individuals aged 16 to 90. It consists of 15 subtests, namely Block Design (BD), Similarity (SI), Digit Span (DS), Matrix Reasoning (MR), Vocabulary (VC), Arithmetic (AR), Symbol Search (SS), Visual Puzzle (VP), Information (IN), Coding (CD), Letter Number Sequencing (LN), Figure Weights (FW), Comprehension (CO), Cancellation (CA), and Picture Completion (PC). The WAIS-IV subtests are grouped into core and supplemental subtests. The first ten subtests are the core subtests and the last five (i.e. LN, FW, CO, CA, and PC) are the supplemental subtests (Wechsler, 2008). The total battery provides an assessment of general intellectual functioning (FSIQ) and four index scores. The four index scales include Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed (Wechsler, 2008b). The Verbal Comprehension scale includes three core subtests (SI, VC and IN) and one supplemental subtest (CO). The Perceptual Reasoning scale includes three core subtests (BD, MR, and VP) and two supplemental subtests (FW and PC). The Working Memory scale includes two core subtests (DS and AR) and one supplemental subtest (LN). The Processing Speed scale includes two core subtests (SS and CD) and one supplemental subtest (CA).

2. Method

2.1. Participants

The Indonesian population size was estimated at 237.6 million in 2010. The majority of people live on the island of Java, and the country's six biggest islands are Kalimantan, Maluku-Papua, Sumatera, Sulawesi, Java, and Bali-Nusa Tenggara (Badan Pusat Statistik Indonesia, 2012). The official language is Indonesian, a standardized register of Malay, an Austronesian language (Lewis, Simons, & Fennig, 2013). Indonesia is a multi-ethnic society,

with more than 1000 ethnic/sub-ethnic groups. However, the size of most ethnic groups is small, and only 15 groups consist of more than one million people (Suryadinata, Arifin, & Ananta, 2003). Formal education and national media are in the Indonesian language (Lewis, Simons, & Fennig, 2013). In addition to being fluent in the Indonesian language, many Indonesians are also fluent in another regional language, such as Javanese, Sundanese, or others. These regional languages are commonly used at home or in the local community.

The participants for this study were 176 Indonesian individuals aged 16 - 75 years old. Of these, 148 participants completed all 15 subtests. Half of the participants (53.4%) were men and 46.6% were women. Age ranged from 16.2 to 83.9 years ($M = 37.34$, $SD = 16.75$). 17.6% of the participants had less than eight years of education, 12.8% completed junior high school, 33.8% senior high school, 7.4% completed academy, 20.3% had an undergraduate degree, 6.8% a master degree and 1.4% a doctoral degree. With respect to ethnicity, participants in our study belonged to the ethnic groups of Tionghoa (37.2%), Java (25%), and Sunda (14.9%). For validity purposes we also recruited participants originally from Ambon, Batak, Betawi, Lampung, Manado, Minang, Nusa Tenggara, Palembang, Papua, Serang, Toraja, and other islands, who lived on Java in the regions mentioned above. Most of the participants were employed (64.86%) in various sectors from education, sanitation, security, transportation, and business sectors.

2.2. Sampling Method

Recruitment of participants was performed in accordance with the WAIS-IV Technical and Interpretive Manual (Wechsler, 2008b), taking the Indonesian life expectancy into account. Based on 1980-2010 Population Census (Badan Pusat Statistik Indonesia, 2012), the life expectancy for Indonesian was 70.7 years. We used a quota sampling method. A stratified sampling plan ensured that the normative sample included representative proportions of individuals according to selected demographic variables. The allocation of samples will be based on the following variables: age, sex, education level, and geographic region. In this study the geographical area covered only Java Island, in and around cities of Jakarta, Tangerang, Bogor, and Bandung.

The involvement of the participants in this research is based on several considerations, as stated in the technical manual of the US WAIS-IV (Wechsler, 2008b). Participants were excluded when their primary language was not Indonesian, when they primarily used nonverbal language or were uncommunicative, when they were unable to understand instructions and participate fully in testing, when they were inadequate to comply with testing to ensure a valid assessment, when they were already tested with any intelligence test in the past six months, when they had an uncorrected visual impairment, uncorrected hearing loss. Upper-extremity disability that would affect motor performance, currently admitted to psychiatric facility, currently taking medication that might impact cognitive test performance (e.g., antidepressants, antipsychotics, etc.), recent and significant change related to cognitive status change, currently receiving chemotherapy in the past two months, and previously diagnosed with any physical condition or illness that might affect to lower the test performance, were excluded from participation.

2.3. Instruments

The original US version of the WAIS-IV was translated into Indonesian by two translators, with bilingual (English-Indonesian) ability. Both translators were educational psychologists and lecturers in the Faculty of Psychology of Atma Jaya Catholic University of Indonesia. They performed the translation in the same period of time. The first two authors (CS, MSH) compared their translations and met with the translators to review the translations, identified differences in Indonesian-English meaning, and adapted the Indonesian-language version to achieve the most accurate culturally equivalent meaning (International Test Commission, 2010).

For the back-translation process, we asked two new independent translators who had bilingual ability and a background in psychology. The first translator was a clinical psychologist and lecturer in the Faculty of Psychology of Atma Jaya Catholic University of Indonesia. The second translator was an organizational psychologist who works as a consultant. Both independently translated the Indonesian WAIS-IV back into US English. Subsequently, the authors and the translators met to review the back translations, identified differences in English-Indonesian meaning, and adapted the Indonesian-language version to achieve the most accurate culturally equivalent meaning. This final translated WAIS-IV was authorized by Pearson Assessment.

The scoring of the WAIS-IV followed the discontinue rules that have been incorporated into the scale to avoid frustration by the examinee because it includes too difficult items. Subtest items were ordered according to

increasing difficulty. The discontinue rules were applied after the examinee had several number of consecutive scores of zero. The discontinue rules are different in each subtest (Wechsler, 2008b). In agreement with Wechsler (2008a), the administration of BD was discontinued when the examinee obtained two consecutive response score of zero. The administration of SI, MR, VC, AR, VP, IN, FW, CO was discontinued when the examinee obtained three consecutive response score of zero. The administration of DS was discontinued when the examinee's response on both trials of one item was zero. The administration of LN was discontinued after the examinee obtained a score of zero on all trials. The administration of PC was discontinued after the examinee obtained four consecutive zero responses. The SS and CD were discontinued after 120 seconds. For CA, each item was discontinued after 45 seconds.

2.4. Procedure

Testing took place in a counseling room at the university, at a company, school, or another place which met the requirements of a psychological testing environment. That is, no external distracting, quiet or soundproof, without interruptions during test administration.

We contacted potential participants and gave information of this study. If a potential participant agreed to participate, we set an appointment to administer the test. Before the test was administered, all participants provided written informed consents. The WAIS-IV was administered in its prescribed order and demographic data were collected. All items in all subtests were administered, and discontinue rules were applied during scoring.

2.5. Analyses

Index difficulty was computed for each item in 12 subtests of the WAIS-IV. SS, CD, and CA were excluded because these are timed tests, for which the index difficulty analyses cannot be performed. In the timed test category, items are so easy that all participants will have all items correct if they have enough time to finish them (Anastasi & Urbina, 1997). As, index difficulty can be different for participants from specific age groups, the next analysis explored whether differences existed between two age groups. The first group consisted of people aged 16 to 69 years and the other of people over 70. This division was based on the administration guidelines of the WAIS-IV subtests, as all subtests are only administered in people from 16 - 69 and only 12 subtests administered in people over 70. We performed this analysis with the Wilcoxon Signed Rank Test.

The next step was to reorder the item sequence, apply the discontinue rules, and calculate the total score for each subtests. We compared these to the original sequence's raw score if we applied the discontinue rules using a dependent sample t-test. Also, we calculated Cohen's *d* (Cohen, 1988). This analysis determined whether the new sequence was the best solution for a subtest.

Reliability coefficients were obtained using Cronbach's coefficient alpha. The alpha coefficient takes the variance of both the subtest total score and item scores into account, and provides the reliability that is the average of all possible split-half reliabilities (Anastasi & Urbina, 1997). Cronbach's coefficient alpha cannot be determined for the subtests of the Processing Speed index (SS, CD, and CA). For the Verbal Comprehension subtests (SI, VC, and CO), we calculated the inter-rater reliability. These three subtests have a different scoring system (Wechsler, 2008a), the criteria for which were also translated into Indonesian. We asked three independent raters to score the participant's verbal responses based on the criteria in the Indonesian manual. Two of three raters were educational psychologists and one a psychometrician. None of the raters had any previous experience with the WAIS-IV scoring rules. Reliability coefficients were obtained using intraclass correlation coefficients (two-way mixed model with absolute agreement).

3. Results

3.1. Item Analysis

For most subtests of the Indonesian translated version of the WAIS-IV, the item index difficulties did not follow the US sequences from easy to difficult (Table 1). Most changes in item sequences had to be made for the subtests in the Verbal Comprehension and Perceptual Reasoning scale (except for BD). Based on this analysis, we re-sequenced the item order of those subtests. Table 1 shows the range and mean of index difficulties across subtests and age categories.

Table 1. WAIS-IV subtest index difficulties.

| Subtest | Range of p for all ages | Mean p for all ages | Mean p for age 16 - 69 | Mean p for age over 70 | Z over age 70 and age 16 - 69 |
|---------|---------------------------|-----------------------|--------------------------|--------------------------|-------------------------------|
| BD | .16 - 1 | .74 | .75 | .70 | -1.14 |
| SI | .05 - .92 | .53 | .53 | .54 | -.078 |
| DS | .02 - 1 | .55 | .55 | .47 | -4.43** |
| MR | .12 - .99 | .57 | .58 | .46 | -3.39** |
| VC | .12 - 1 | .58 | .58 | .58 | -.080 |
| AR | .14 - .99 | .56 | .56 | .56 | -.260 |
| VP | .11 - 1 | .53 | .53 | .45 | -2.98** |
| IN | .02 - .99 | .33 | .33 | .33 | -.24 |
| LN | .04 - .99 | .62 | .62 | n/a | n/a |
| FW | .12 - .97 | .50 | .50 | n/a | n/a |
| CO | .00 - .82 | .34 | .34 | .37 | -1.047 |
| PC | .02 - .96 | .37 | .38 | .28 | -3.30** |

**Wilcoxon signed rank test (Z test) is significant at the .01 level (2-tailed).

Subsequently, we analyzed the differences in index difficulties between the two age groups. Using the Wilcoxon Signed Rank Test, Z-test, we found that there were significant results on four subtests (DS: $Z = -4.43$, $p < .01$; PM: $Z = -3.39$, $p < .01$; VP: $Z = -2.98$, $p < .01$; PC: $Z = -3.30$, $p < .01$). All these subtests were more difficult for participants over age 70 years.

After rearranging the item order, we scored the performance using the discontinue rules (Table 2). Significant differences in performances between the original and rearranged sequences were found for VC ($t_{(147)} = -5.01$, $p < .01$), VP ($t_{(147)} = -3.60$, $p < .01$), IN ($t_{(147)} = -6.50$, $p < .01$), and CO ($t_{(147)} = -12.81$, $p < .01$). Of these, CO had the largest effect size. No significant differences were found for BD, SI, MR, AR, FW, and PC.

3.2. Reliability Analysis

Table 3 shows the reliability coefficients of each WAIS-IV subtests. Cronbach's coefficient alpha of the WAIS-IV subtest ranged from .74 - .92, indicating an acceptable (BD, SI, and CO) to excellent (LN and FW) reliability. For the subtests of the Verbal Comprehension index, we also computed inter-rater reliabilities. The inter-rater agreements were high for SI ($r = .97$), VC ($r = .96$), and CO ($r = .91$).

3.3. Demographic Data Analysis

We found no significant differences in the scores of men and women in any of the subtests. Significant differences between participants of different educational backgrounds were found. Table 4 shows the differences between participants that had completed senior high school ($N = 50$) and those who had a university undergraduate degree ($N = 30$).

Significant differences were found on most subtests, except CO ($t_{(78)} = .22$, $p > .05$) and CA ($t_{(78)} = 1.47$, $p > .05$). The group that had an undergraduate degree scored higher on all subtests, in agreement with previous research (Matarazzo & Herman, 1984; Grieve & Van Eeden, 2010).

Table 5 shows the Pearson product-moment correlation coefficients between age and all subtests in WAIS-IV. We found significant correlations for all subtests of the Processing Speed index, most of the subtests of the Perceptual Reasoning index, and most of the subtest of the Working Memory index. No significant correlations were found between age and the subtests of the Verbal Comprehension index.

For the subtests of the Processing Speed index, the highest correlation was between age and SS ($r_{(146)} = -.40$, $p < .01$). These results were consistent with numerous studies about aging-related decline in speed of information processing (see Verhaeghen & Salthouse, 1997). FW was the only subtest from the Perceptual Reasoning index that did not significantly correlated with age. The highest correlation was found between age and VP, $r_{(146)} = -.25$, $p < .01$. These results were consistent with studies about age and reasoning (see Verhaeghen & Salthouse, 1997). For the Working Memory index, significant correlations were found between

Table 2. The differences between original and rearrange sequence of subtests' items, scored with discontinue rules (df = 147).

| Subtests | Mean score of | | t-value | Sig. (2 tailed) | Cohen's d |
|---------------|---------------|------------|---------|--------------------|-----------|
| | Original | Rearranged | | | |
| BD-BD_Reorder | 34.35 | 34.40 | -1.42 | .16 | .12 |
| SI-SI_Reorder | 18.57 | 18.61 | -.29 | .77 | .02 |
| MR-MR_Reorder | 13.13 | 13.38 | -1.67 | .10 | .14 |
| VC-VC_Reorder | 30.20 | 33.09 | -5.01 | .00 | .41 |
| AR-AR_Reorder | 11.90 | 11.92 | -.24 | .81 | .02 |
| VP-VP_Reorder | 11.66 | 12.24 | -3.60 | .00 | .30 |
| IN-IN_Reorder | 7.02 | 7.91 | -6.50 | .00 | .53 |
| FW-FW_Reorder | 11.74 | 11.86 | -.98 | .33 | .08 |
| CO-CO_Reorder | 4.88 | 11.84 | -12.81 | .00 | 1.05 |
| PC-PC_Reorder | 8.22 | 8.32 | -1.22 | .22 | .10 |

Table 3. The subtests' reliability coefficients.

| Name of the subtest | Alpha | SEM | N items | Mean | SD |
|-------------------------------|-------|------|---------|-------|-------|
| Block Design (BD) | .77 | 5.80 | 14 | 35.29 | 12.12 |
| Similarity (SI) | .79 | 2.56 | 18 | 19.19 | 5.53 |
| Digit Span (DS) | .89 | 2.19 | 48 | 26.16 | 6.67 |
| Matrix Reasoning (MR) | .87 | 1.96 | 26 | 14.93 | 5.40 |
| Vocabulary (VC) | .87 | 3.56 | 30 | 34.99 | 9.84 |
| Arithmetic (AR) | .86 | 1.58 | 22 | 12.34 | 4.30 |
| Symbol Search (SS) | n/a | n/a | n/a | 28.14 | 1.37 |
| Visual Puzzle (VP) | .80 | 1.91 | 26 | 13.69 | 4.24 |
| Information (IN) | .88 | 1.59 | 26 | 8.43 | 4.68 |
| Coding (CD) | n/a | n/a | n/a | 59.80 | 23.15 |
| Letter Number Sequencing (LN) | .92 | 1.66 | 30 | 17.43 | 6.04 |
| Figure Weights (FW) | .90 | 1.97 | 27 | 13.48 | 6.17 |
| Comprehension (CO) | .74 | 2.61 | 18 | 12.39 | 5.16 |
| Cancellation (CA) | n/a | n/a | n/a | 3.67 | 12.15 |
| Picture Completion (PC) | .88 | 1.74 | 24 | 8.95 | 4.92 |

DS ($r_{(146)} = -.30, p < .01$) and LN ($r_{(137)} = -.39, p < .01$). These results are consistent with studies about age and working memory (see Verhaeghen & Salthouse, 1997).

4. Discussion

In this study, we described the translation process of the WAIS-IV into the Indonesian language. We performed psychometric analysis, including the order of the item sequences, item analyses, and reliability of each subtests of the WAIS-IV-ID.

As expected, the item sequences of most of the translated WAIS-IV subtests had to undergo major changes. In all subtests from the Verbal Comprehension index, we had to reorder the item sequence, although stayed close to the original items for content purposes. For example, in the subtest information of the original US test, most of the items that refer to science, geography, world history, world figures, and literature are related to Western culture knowledge. Indonesian participants were more likely to answer correctly in science and geography, but most of them had difficulties in answering items about specific historical persons (like Sacagawea) and literature (Alice in the Wonderland). The index difficulties ranged from .41 - .02. On those items, participants either gave up or said they remembered that this figures had appeared in a movie. In one item about historical figures in the US, their answers were not fully correct although they gave correct information to some extent. Another example is the VC subtest, which also had to undergo major reordering, because the items may have different meaning

Table 4. Subtests differences between groups with different education backgrounds (df = 78).

| | M for high school | SD for high school | M for university | SD for university | t-test | Sig (2-tailed) | Cohen's d |
|----|-------------------|--------------------|------------------|-------------------|--------|----------------|-----------|
| BD | 34.46 | 11.50 | 44.20 | 11.12 | 3.71 | .00 | 3.71 |
| SI | 19.42 | 4.41 | 22.57 | 3.77 | 3.26 | .00 | 3.26 |
| DS | 26.30 | 6.41 | 30.27 | 5.63 | 2.80 | .01 | 2.80 |
| MR | 14.72 | 5.21 | 18.80 | 4.25 | 3.62 | .00 | 3.62 |
| VC | 34.12 | 8.14 | 41.57 | 7.00 | 4.17 | .00 | 4.17 |
| AR | 12.14 | 4.66 | 15.07 | 3.47 | 2.98 | .00 | 2.98 |
| SS | 28.98 | 9.79 | 34.30 | 10.45 | 2.29 | .02 | 2.29 |
| VP | 13.76 | 3.55 | 17.07 | 4.06 | 3.82 | .00 | 3.82 |
| IN | 8.00 | 3.61 | 11.73 | 4.16 | 4.23 | .00 | 4.23 |
| CD | 61.18 | 19.81 | 76.33 | 17.39 | 3.46 | .00 | 3.46 |
| LN | 17.60 | 5.83 | 20.47 | 5.04 | 2.24 | .03 | 2.24 |
| FW | 13.88 | 6.20 | 17.00 | 5.61 | 2.26 | .03 | 2.26 |
| CO | 13.68 | 4.43 | 13.90 | 4.16 | 0.22 | .83 | 0.22 |
| CA | 31.76 | 11.40 | 35.90 | 13.53 | 1.47 | .15 | 1.47 |
| PC | 8.70 | 4.74 | 12.63 | 4.31 | 3.71 | .00 | 3.71 |

Table 5. The subtests' correlation with participant's age.

| | BD | SI | DS | MR | VC | AR | SS | VP | IN | CD | LN | FW | CO | CA | PC |
|-----------------|--------|-----|--------|-------|-----|------|--------|--------|------|--------|--------|------|-----|--------|--------|
| Age | -.24** | .08 | -.30** | -.18* | .06 | -.02 | -.40** | -.25** | -.01 | -.39** | -.39** | -.16 | .05 | -.36** | -.22** |
| r ² | .06 | .01 | .09 | .03 | .00 | .00 | .16 | .06 | .00 | .16 | .15 | .03 | .00 | .13 | .05 |
| Sig. (2-tailed) | .00 | .31 | .00 | .03 | .48 | .83 | .00 | .00 | .88 | .00 | .00 | .06 | .53 | .00 | .01 |

*Correlation is significant at the .05 level (2-tailed); **Correlation is significant at the .01 level (2-tailed).

in daily conversations compared to their formal meaning in Indonesian. For instance, item number 4 (“bed”, in Indonesian “ranjang”) of VC, became item 11 after reordering. Specific syllables in some words in Indonesia, like the “ke” can be “erased” when spoken. In item number four, the syllable “ke” is not present. The participants who gave incorrect answers in this item, tended to listen to the word but not read it. Even if the examiner had pointed to the written word, many participants still not changed their answer. As a result, many zero responses were scored for this item and its index difficulty was .76. Another example is item number 17 (i.e. “plagiarism”, in Indonesian “plagiat”), after reordering, it became item 30. The subtest Arithmetic (AR) from Working Memory index and all subtests from Perceptual Reasoning index, except for Block Design, also needed major reordering of the items sequence. Many of the Picture Completion items resulted in incorrect responses from the participants, because they did not indicate the important missing detail of the picture, but indicated other absent objects in the picture instead. For instance, for the item in which a picture with trees is shown, participants answered that the river lacked fish, or that people or vehicles were missing on the picture. Other incorrect responses were found in several pictures with which Indonesian people are unfamiliar. For example, items such as snow or a stove are less common in Indonesia (many participants responded that the stove is a kind of washing machine). Clearly, reordering the items and applying the discontinue rules improved the participants' scores.

Some other recommendations can be made based on our experiences in the data collection process. First, in the similarity subtest, the example item (what is the similarity between “two and seven”) was difficult to answer for some participants. They answered in a concrete way, e.g. “both have edges”. If participants do not understand the instruction, then continuation is a problem. To overcome this, we repeated the instruction twice and explained the correct answer of the example item. In case they still did not understood the purpose of the subtest, the examiner gave an example using two concrete objects, e.g. what is the similarity of two pieces of fruit (mango and banana), that are more part of the participants' daily lives. We used this procedure especially in participants with a lower educational background and for senior citizens. Secondly, in the Comprehension subtest we found no significant differences between the performances of participants who completed high school and

university graduates. This result was not consistent with other studies (Matarazzo & Herman, 1984; Grieve & Van Eeden, 2010). By checking the responses of those participants, we found that many of the answers were very short. From the data collection evaluation, the participants tended to say, “Ok, that’s it” or “I don’t know anything else” or “It’s enough”. Their responses usually only covered one general concept, for which they did not obtain the maximum two points. For future research, we recommend to rephrase the item appropriately to obtain another response as suggested in the Wechsler (2008a).

We found age effects that are in agreement with previous research. For instance, Verhaeghen and Salthouse (1997) performed a meta-analysis on 50 studies about aging and speed of processing, in which they reported correlations between $-.23$ and $-.68$. In our study, we found correlations between $-.36$ and $-.40$. Verhaeghen and Salthouse (1997) also performed a meta-analysis on 34 studies in which the relation between age and the Working Memory index was studied. They found that correlations varied between $.03$ and $-.48$. In our study, we found a range between $-.02$ until $-.39$. These negative correlation coefficients are consistent with previous studies, and indicate that the number of items that could be remembered immediately decreases with age and affects the total score. Overall, most of the correlations between WAIS-IV subtests with age, showed significant and negative correlations. Verhaeghen and Salthouse (1997) suggest that almost all aspects and type of information processing are affected by age, causing a broad decline in many facets of cognitive functioning when people get older. They concluded that there are two types of general factors that explain the cognition differences in adult. The first type refers to basic and relatively pervasive loss in processing speed. The second type refers to the ability to preserve information in a temporary working memory store while processing is carried out. These two types are not independent because the proportion of age related variance in cognitive performance that is related to working memory capacity is also shared to large extent with processing speed measures. Like in their study our results also show that the subtests from the Processing Speed index and most of Working Memory index are correlated with age. We did not find this correlation for AR from the Working Memory index, because this subtest also depends on the more crystallized verbal and quantitative capacities.

Comparing the reliability coefficients from the Indonesian adaptation of the WAIS-IV to the US version (Wechsler, 2008b), all reliability coefficients are acceptable to good (although in general somewhat lower than the US coefficients). Moreover, the subtests in the Verbal Comprehension index have an excellent inter-rater agreement reliability. A possible weakness is the probing used by examiner. We tested participants from different educational and ethnic backgrounds. Some of the participants needed explanations in their regional language, such as Javanese, Sundanese, or Mandarin. Sometimes the participants gave mixed responses in regional language and Indonesian language, but the examiner sometimes was less familiar with the regional language. For further research, we recommend to recruit examiners from different ethnic backgrounds who understand the regional language. Moreover, recruitment from this study was limited to Jakarta and its surroundings. For further research, the geographical area has to be expanding to other areas of Indonesia. Also, a larger sample is recommended to replicate the findings of this study and to analyze the construct validity of the translated version.

Overall, the results of this first study provide a strong foundation for developing the Indonesian version of WAIS-IV further. It is psychometrically promising; the reliabilities of the subtests are generally good, and have a high inter-rater agreement. The next step is to explore the factorial structure of the Indonesian WAIS-IV using a larger sample.

Acknowledgements

This research is supported by research grants from Indonesian Ministry of Education and Culture under the International Research Collaboration and Scientific Publication in 2013 (No. 085/SP2H/PL/DIT.LITABMAS/V/2013). The authors acknowledge the contribution of Darma Juwono, M.Psi and Dr. Weny Sembiring, for their work on translations. Our gratitude goes to Justinus Budi Santoso, M.Psi and Rendy Wirawan, M.Psi, for their work on back translations and their input in the adaptations. The authors are very grateful to Isabella, Stella Rosalina, Widhi Adhiatma, Stephanie, Yohana Natalia Tryastuti, Amelia Yulistiani, Justinus Budi Santoso, Stephanie Nessia, Cindy Komara, Jessica Emily, Yulianti Sanjaya, and JessicaTania for their contribution in data collection of this research. Copyright © 2008 NCS Pearson, Inc. Indonesian translation copyright © 2013. Translated, adapted, and reproduced with permission of publisher. All rights reserved. We thank Pearson Assessment for their approval.

References

- Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Badan Pusat Statistik Indonesia (2012). Perkembangan beberapa indikator utama sosial-ekonomi Indonesia [Trends of Selected Socio-Economic Indicators of Indonesia]. http://www.bps.go.id/booklet/Booklet_Agustus_2012.pdf
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Fitrikasari, A., Jaeri, S., & Bintoro, A. C. (2013). Comprehensive Care of Epilepsy: Involvement of Psychological Assessment and Wada Test. *Basic Research Journal of Medicine and Clinical Sciences*, 2, 1-4.
- Grieve, K. W., & Van Eeden, R. (2010). A Preliminary Investigation of the Suitability of the WAIS-III for Afrikaans-Speaking South Africans. *South African Journal of Psychology*, 40, 262-271. <http://dx.doi.org/10.1177/008124631004000305>
- International Test Commission (2010). International Test Commission Guidelines for Translating and Adapting Tests. <http://www.intestcom.org>
- Lange, R. T. (2007). WAIS-III Index Score Profiles in the Canadian Standardization Sample. *Journal of Clinical and Experimental Neuropsychology*, 29, 47-58. <http://dx.doi.org/10.1080/13803390500409591>
- Lecerf, T., Golay, P., & Reverte, I. (2012). CHC Composite Scores for the WAIS-IV: French Norms. *Pratiques Psychologiques*, 18, 401-412. <http://dx.doi.org/10.1016/j.prps.2012.03.001>
- Lembaga Pengembangan Sarana Pengukuran dan Pendidikan Psikologi Fakultas Psikologi Universitas Indonesia (n.d.). *Manual Wechsler-Bellevue Intelligence Scale for Adolescents and Adult Form I*. Depok: Limited Used in Psychology Profession.
- Lewis, M. P., Simons, G. F., & Fennig, C. D. (Eds.) (2013). *Ethnologue: Languages of the World* (17th ed.). Dallas, TX: SIL International.
- Lynn, R., & Dai, X. Y. (1993). Sex Differences on the Chinese Standardization Sample of the WAIS-R. *The Journal of Genetic Psychology: Research and Theory on Human Development*, 154, 459-463. <http://dx.doi.org/10.1080/00221325.1993.9914744>
- Matarazzo, J. D., & Herman, D. O. (1984). Relationship of Education and IQ in the WAIS-R Standardization Sample. *Journal of Consulting and Clinical Psychology*, 52, 631-634.
- Melendez, F. (1994). The Spanish Version of the WAIS: Some Ethical Considerations. *Clinical Neuropsychologist*, 8, 388-393. <http://dx.doi.org/10.1080/13854049408402041>
- Murayama, N., Iseki, E., Tagaya, H., Ota, K., Kanasuki, K., Fujishiro, H., Arai, H., & Sato, K. (2013). Intelligence or Years of Education: Which Is Better Correlated with Memory Function in Normal Elderly Japanese Subjects? *Psychogeriatrics*, 13, 9-16. <http://dx.doi.org/10.1111/j.1479-8301.2012.00408.x>
- Polhaupessy, L. F. (n.d). Tes Wechsler-Bellevue. Bandung: Internal Used Publication.
- Roivainen, E. (2013). Are Cross-National Differences in IQ Profiles Stable? A Comparison of Finnish and U.S. WAIS Norms. *International Journal of Testing*, 13, 140-151. <http://dx.doi.org/10.1080/15305058.2012.657315>
- Seksi, P. (1992). *Buku Pegangan Wechsler Adult Intelligence Scale*. Yogyakarta: Fakultas Psikologi Universitas Gadjah Mada.
- Suryadinata, L., Arifin, E. N., & Ananta, A. (2003). *Indonesia's Population: Ethnicity and Religion in a Changing Political Landscape*. Singapore City: Institute of Southeast Asian Studies.
- Van der Heijden, P., Van den Bos, P., Mol, B., & Kessels, R. P. C. (2013). Structural Validity of the Dutch-Language Version of the WAIS-III in a Psychiatric Sample. *Applied Neuropsychology: Adult*, 20, 41-46. <http://dx.doi.org/10.1080/09084282.2012.670166>
- Verhaeghen, P., & Salthouse, T. A. (1997). Meta-Analyses of Age-Cognition Relations in Adulthood: Estimates of Linear and Nonlinear Age Effects and Structural Models. *Psychological Bulletin*, 122, 231-249.
- Weiss, L. G., Saklofske, D. H., Coalson, D., & Raiford, S. E. (Eds.) (2010). *WAIS-IV Clinical Use and Interpretation*. San Diego, CA: Academic Press.
- Wechsler, D. (2008a). *Wechsler Adult Intelligence Scale—Fourth Edition Administration and Scoring Manual*. San Antonio, TX: Pearson.
- Wechsler, D. (2008b). *Wechsler Adult Intelligence Scale—Fourth Edition Technical and Interpretive Manual*. San Antonio, TX: Pearson.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or [Online Submission Portal](#).

