Cross-cultural applicability and reduction of the American seven-subtest short form of the WAIS on a Swedish non-clinical sample

LINDAU MARIA & NAJSTRÖM MATS

Corresponding address: Maria Lindau, Department of Psychology, Stockholm University, Stockholm SE-106 91, Sweden. Email: maria.lindau@psychology.su.se

Abstract

The study aimed at investigating whether the seven-subtest short form based on WAIS-R (Ward 1990) was statistically valid to use on the Swedish version of Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV), if this abbreviation was fit to catch the heterogeneity in test performance across age and if this brief measure was possible to abbreviate even more. WAIS-IV data from a non-clinical sample consisting of 261 participants ranging between 18 and 74 in age was analyzed with bivariate and multiple regression analyses, a prorating method for calculation of Full Scale IQ (FSIQ) and its indices as well as paired-samples t-test. The results were contradictory. When the original WAIS-IV was compared to the seven-subtest short form the results showed a good congruence on FSIQ-level between the two sets, but on index level there were several cases of mismatches. In the younger and middle aged sample (<55 years) results on FSIQ as well as index level were in accordance, whereas in the elderly group (~55 years) they were incongruent. The best reduction of the seven-subtest short form was a four-subtest model, encompassing Block Design, Similarities, Arithmetic and Coding, one subtest from each index, but the t-tests indicated several cases of mismatches between the full WAIS-IV measures and the prorated scores. Applied on the Swedish version of the WAIS-IV the seven-subtest formula appears to be applicable on an FSIQ level, to be suitable for a younger sample, but not for an elderly. Otherwise, this model and the four-subtest model are recommended to be used with caution.

Keywords: WAIS-IV, short form, cross-cultural, Scandinavian version, age heterogeneity

Department of Psychology, Stockholm University, Stockholm SE-106 91, Sweden

Preliminary versions of this study on two sets of smaller samples have earlier been presented at the 4th Global Experts Meeting on Neuropharmacology, September 14–16, 2016, San Antonio, USA, and 8th Annual Pharma Middle East Congress during October 10–12, 2016 in Dubai, United Arab Emirates (UAE).

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/Licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
Introduction

One of the internationally most frequently applied neuropsychological test batteries is the Wechsler Adult Intelligence Scale (WAIS) (Ryan, Kreiner, Gontkovsky, & Glass Umfleet, 2015). It has also been described as the “work horse” in clinical assessments (Crawford, Allum, & Kinion, 2008). The WAIS originates in the Wechsler-Bellevue Intelligence Scale (1939), revised and modernized into the following versions: the Wechsler-Bellevue Form II (1946), Wechsler Ault Intelligence Scale (1955), Wechsler Ault Intelligence Scale – Revised (WAIS-R, 1981), WAIS-III (1997) and WAIS-IV (2008), which is the most recent update of the scale.

The test framework of the WAIS-IV provides composite scores that represent intellectual functioning in four specific cognitive areas: Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI), and Processing Speed Index (PSI). Each index scale contributes to a composite score that represents general intellectual ability, the Full Scale IQ (FSIQ). The indices are composed by sets of subtests identified as core or supplemental. The VCI includes three core subtests: Similarities, Vocabulary, Information and Comprehension as a supplemental subtest. Accordingly, PRI includes three core subtests: Block Design, Matrix Reasoning, Visual Puzzles, together with Figure Weights and Picture Completion as supplemental subtests. The WMI includes two core subtests: Digit Span, Arithmetic and Letter-Number Sequencing as a supplemental subtest. Finally, PSI comprises two core subtests: Symbol Search, Coding, and Cancellation as a supplemental subtest.

Estimating FSIQ and mapping the patient’s entire cognitive profile is a time-consuming procedure. The time required to implement the full-scale WAIS vary significantly depending on whether the time estimates concern the assessment of normative samples or clinical patients. According to the WAIS-IV manual (Wechsler, 2008) the administration time for the ten core tests varies between 100 minutes for 95% of the standardization sample and 59 minutes for 95% of the category intellectually disabled. The awareness of the long time required to make cognitive assessment dates back a century (Silverstein, 1990) and up to now have a multitude of abbreviated Wechsler scales been published. For example, recently Úbeda, Fuentes, and Dasí (2016) recommended the use of short forms when the objective is to assess global intellectual functioning.

The number of subtests in proposed abbreviated measures varies between two and eight and, the time gain is of course depending on which and the number of subtests included. An abbreviation of the Wechsler scale published in 2011, the Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II), comprises two as well as four subtests, with the subtests Vocabulary and Matrix Reasoning composing the dyad and Vocabulary, Similarities, Block Design and Matrix Reasoning creating the four-subtest form. Completion time of the two sets of tests is estimated to 15 for the dyad and 30 minutes for the quadruple (McCrimmon & Smith, 2013). Both these short forms enable estimations of FSIQ. Unfortunately, none of these abbreviations includes subtests from the remaining indices, that is WMI and PSI, which is important to be able to give a picture of the global cognitive ability. To our knowledge there is no Swedish standardization of the WASI-II. On a German-speaking sample, the efficiency of the WASI-II dyad was however supported by findings by Daseking, Petermann, and Waldmann (2014), who argue for a selection of the two subtests Vocabulary and Matrix Reasoning, as a viable short form of the WAIS-IV in specific clinical settings. Other short forms are developed by Denney, Ringe, and Lacritz (2015) who suggest different dyadic short forms
for estimating FSIQ, with results showing a strong association with actual FSIQ (rs = .73 to .88). This ambition to make as extensive abbreviations of the WAIS as possible is in line with findings by Girard, Axelrod, Patel, and Crawford (2015) who have identified several dyadic combinations subtests predicting FSIQ with high reliability and validity. The Coding and Information dyad was the strongest predictor with the composite value $R_c = .887$. However, as the authors suggest, dyads should be used with caution, since their applicability depends on the actual referral or research question. More extensive short forms have been evaluated by Girard, Axelrod, and Wilkins (2010). In a mixed clinical sample, the authors demonstrated the feasibility of WAIS-III abbreviations consisting of seven as well as eight subtests. The authors concluded that all three tested short forms reduced testing time but otherwise had different strengths. For example, the four factors, i.e. the indices, were best represented by an eight-subtest short form whereas the Ward (1990) seven-subtest short form included in the study (Girard et al., 2010) had the best psychometric qualities. This seven-subtest short form, originating in the WAIS-R had correlations of .96, .97, and .98 with the Performance IQ (PIQ), Verbal IQ (VIQ), and Full-Scale IQ (FSIQ), respectively, of their full-length WAIS-R equivalents (Ward 1990) and is one of the currently most applied abbreviations (e.g. Bulzacka et al., 2016; Strauss, Sherman, & Spreen, 2006). Thus, in the literature much speaks for the benefit of the Ward (1990) formula. However, this seven-subtest short form emanates from an American standardization, and as far as we know, there exists no investigation of the applicability of it in a Swedish context. Because of this lack of normative studies of the short form in a Swedish context, the main focus of the present study was to fill this knowledge gap.

The seven-subtest short form encompasses the following seven subtests: Information, Digit Span, Arithmetic, Similarities, Picture Completion, Block Design and Coding (in WAIS-R called Digit Symbol). In the WAIS-IV all the subtests in the short form according to Ward (1990) are core tests, except for Picture Completion, which in the WAIS-IV is a supplemental test that can be used as a substitution for one of the core subtests Block Design, Matrix Reasoning or Visual Puzzles. The validity of a shortening of WAIS-IV according to the Ward (1990) formula applied to clinical patients has been demonstrated by Meyers, Zellinger, Kockler, Wagner, and Miller (2013). The authors investigated the applicability of the seven-subtest short form by creating a regression equation based on one dataset, validating these data on a second dataset, and finally comparing the results according to the prorated score method from the WAIS-IV manual. The results showed a strong correlation between the methods which can be applied interchangeably in clinical settings. This is also in line with an earlier study by Pilgrim, Meyers, Bayless, and Whetstone (1999).

The Ward (1990) short form, like most of the intelligence tests, has been created and standardized in the US, and thereafter “travelled to other countries” (Bowden, Saklofske, & Weiss, 2011a, p. 133; Roivainen, 2013). The prevailing view is that neuropsychological tests must take cultural differences into consideration, with respect to the content of the testing tasks and translation issues (Escobedo & Hollingworth, 2009; Wechsler, 2010) as well as the norms used to estimate the performance levels (Roivainen, 2013; Wechsler, 2010). Regarding the sets of norms, a one-to-one correspondence between the American WAIS-IV national standardization data and performance levels in other cultures is not to be automatically expected. Cultural disparities are also the reason for the existence of e.g. Scandinavian, Dutch, German, French and Spanish adaptations of testing tasks and evaluation norms of the original American version of the WAIS-IV.
The Scandinavian adaptation of the WAIS-IV currently in use was published in 2010. The standardization was conducted between 2008 and 2010, based upon a normative sample stratified according to age, gender and education, defined by Statistics Sweden, and the Danish and Norwegian equivalents, consisting of a total of 726 individuals (364 females and 362 males). This version is from 16:0 years up to 74:11 years based on Scandinavian norms whereas the age span ranging from 75:0 to 90:11 years in the Scandinavian adaptation is based on American norms (Wechsler, 2010). The Scandinavian version provides common norms for the three Nordic countries Sweden, Denmark and Norway, but is translated into the three, different Nordic languages. To ascertain the validity of the Scandinavian adaptation the same statistical analyses as in the American version have been conducted, i.e. factor analyses, inter-correlation analyses, and associations with other measures. The factor structure of the Wechsler tests has been the subject of numerous studies, and there is an ongoing debate whether a four-factor or a five-factor model of the American version of the WAIS-III and the WAIS-IV best describes the latent structure. There are findings supporting a four-factor model as being the most suitable (Bowden, Lissner, McCarthy, Weiss, & Holdnack, 2007; Bowden, Saklofske, & Weiss, 2011b; Wechsler, 2003, 2008). However, a five-factor model has also been suggested as an option with good fit (Ward, Bergman, & Hebert, 2012; Weiss, Keith, Zhu, & Chen, 2013). In the Scandinavian adaptation of the WAIS-IV a four-factor model has been chosen as a basis for score interpretation. The factor loadings for the indices range from .44 to .90 (Wechsler, 2010). The inter-correlations between the subtests within the indices VCI and PRI ranged as follows: VCI: .64 to .70; PRI: .65 to .75. For WMI and PSI the inter-correlations amounted to .50 and .64 respectively. This correlation pattern coincides with patterns emerging in WAIS-III and earlier versions of the Wechsler scale (Wechsler, 1989, 1991, 1997, 2002, 2003). The average reliability coefficients across age for the four indices in the Scandinavian WAIS four-factor model are strong and varies from .90 to .94.

Another aspect of test variability is how well the WAIS is apt to reflect performance levels across age. Ardila (2007) found differences between WAIS-III subtests in how well they captured the variability in intellectual functioning among the elderly. According to Ardila (2007) the proliferation of performance levels increased during normal aging, particularly in Matrix Reasoning, Letter-Number-Sequencing, Digit-Symbol, Picture Completion and Picture Arrangement, where the rise in the spread was more than 200%. In the subtests Block Design, Object Assembly and Information the enlargement of heterogeneity in test scores was considerably lower, <20%. A similar pattern was observed in the WAIS-IV by Wisdom, Mignogna, and Collins (2012), who reported a strong increase, between 56% and 98%, of age-related subtest performance heterogeneity on raw-score level in the subtests Block Design, Matrix Reasoning, Picture Completion, Symbol Search and Coding. In the remaining subtests the variability was lower, ranging from 32% to <14%.

As earlier mentioned, of importance is that the subtests included in the Ward formula is a combination of tests whose validity and reliability are based on norms from an American standardization (Meyers et al., 2013), meaning that its usefulness in Sweden has not been confirmed. The possible limitations in the full WAIS-IV scale concerning cross-cultural transferability of the battery, and irregularities in the subtests’ sensitivity to age related variations in test performance may consequently have been transplanted to the seven-subtest short form. Thus, the aim was to study whether the seven-subtest short form according to Ward (1990) was statistically valid to use on the Swedish version of the WAIS-IV (Wechsler, 2010),
the applicability of the seven-subtest short form on younger and older individuals, and in order to explore the possibilities of a further increase of time gain in administration time, to investigate if, and in that case how, a further reduction of the seven-subtest short form (Ward 1990) would be possible to implement on the Swedish adaptation of WAIS-IV, with preserved good statistical certainty.

**Method**

**Site of the study**
The investigation was conducted at the Department of Psychology, Stockholm University, Sweden.

**Participants**
The WAIS-IV was administered by graduate students in psychology at the Stockholm University, under the supervision of experienced licensed psychologists. The majority of the participants were recruited from the students’ social networks among Swedish speaking, non-clinical participants consenting to participate in the study. Occasionally, participants were recruited from senior housings, where the staff had suggested elderly, healthy persons suitable for neuropsychological testing, and able to give their informed consent to the participation. Sometimes participants were enrolled among persons who had contacted the Department of Psychology on their own initiative and reported their interest in participating in cognitive testing. Participation in the study was ascertained through a pre-testing interview. Participants reporting disturbances related to cognition, neuropsychiatric disorders, depression or other psychiatric conditions, alternatively ongoing treatments with strenuous side effects, such as radio- or chemotherapy were excluded. The majority of the participants completed all the core tests and Picture Completion comprised in the WAIS-IV, however in some cases, participants omitted single subtests in the battery, resulting in a somewhat varied number of participants included in the different analyses. The total sample consisted of 261 individuals ranging in age from 18 to 73 (M = 42.77; SD = 17.73). The main part of the participants had ≥ 12 years of education, corresponding to upper secondary school level or above.

To analyse the predictive power of the seven-subtest short form and its reduction across ages the sample was divided in two age groups, one younger < 55 years (M = 31.79; SD = 9.55) composed by 175 participants and one older ≥ 55 years (M = 65.12; SD = 4.58) including 86 participants. The age 55 was selected as principle of division, since it made it possible to include a relatively large group of elderly participants, and the division corresponds to an age interval in the WAIS-IV norms (Wechsler, 2010).

**Procedure**

**Test administration**
WAIS-IV subtests have been administered and scored according to standardized procedures described in the Scandinavian version of WAIS-IV (Wechsler, 2010).

**Statistics**
The analyses of the test results were based on weighted points: for the subtests scaled score equivalents of the raw scores and for the indices as well as the FSIQ, composite scores. Data was analyzed using bivariate and multiple regressions in IBM SPSS Statistics.
version 24. It is important to acknowledge that the measures of the full scale WAIS-IV and short forms in this study are extracted from the same dataset obtained by a single administration, hence, short forms measures are embedded in the full scale WAIS-IV data. Due to this replicated, non-independent measurement error shared by the part (short forms) and its whole (full scale) the strength of the calculated short forms - full scale correlation may be inflated. The obtained correlations can be misleading if looked at as a typical validity coefficient (Tellegen & Briggs, 1967). In order to broaden the analyses of the correlations between the short forms and full scale WAIS-IV obtained from the regression analyses, a prorated procedure for calculation of FSIQ and indices proposed by Paolo and Ryan (1993) was applied. Prorating derives composites from the average of the selected subtests administered. For example, in the case of the seven-subtest short form calculating index PRI (based on three subtests in full scale WAIS-IV) the procedure is as follows: the sum of the two age-scaled subtests are multiplied by 3/2. This value is referenced in the test manual (Wechsler, 2010; Table A.3–7) to obtain a corresponding index score.

Results

As is seen in Table 1 the mean cognitive performance levels on the indices as well as on the FSIQ were above average in the group. The index values spanned from borderline level, PSI =71, to a very superior level ≥130 on FSIQ and all its indices. The mean cognitive performance levels in all investigated aspects ranged from high to high average (for classification of performance levels, see Chevalier, Stewart, Nelson, McInerney, & Brodie, 2016). Despite the high FSIQ-values, a one sample Kolmogorov-Smirnov test of normality showed that the distribution of scores was non-significant (p >.05) and thus that data was normally distributed.

There was no indication of problematic collinearity levels in any of the models, as all variables possessed a tolerance value >.9 as well as VIF value <10.

To investigate whether the American short form was statistically valid to use on the Swedish version of WAIS-IV, the seven subtests (Ward 1990) were entered as predictors into a multiple regression using the standard method with FSIQ as criterion variable. A significant model emerged: $F(7, 210) = 417.853, \ p <.001$. The model explained 93.1% of the variance in FSIQ (adjusted $r^2 = .931$), see Table 2. All predictor variables contributed significantly to the explained variance in FSIQ.

Table 1. Results on WAIS-IV indices and FSIQ.

| Test | Points \(^a\) | Min | Max | Mean | SD  |
|------|--------------|-----|-----|------|-----|
| VCI  |              | 87  | 150 | 118.97 | 12.25 |
| PRI  |              | 74  | 148 | 113.92 | 14.21 |
| WMI  |              | 79  | 150 | 111.91 | 15.01 |
| PSI  |              | 71  | 147 | 109.02 | 13.20 |
| FSIQ |              | 86  | 143 | 115.70 | 11.63 |

\(^a\)Composite scores.
With the reduction of the full scale to the seven-subtest short form has also followed a cut of the number of subtests included in all the indexes, except WMI which retains its two subtests. In order to test the validity of the different short form index scales for FSIQ, three regression analyses with the index subtests were conducted.

1. The VCI was used as a criterion variable in a multiple regression using the standard procedure, with the two subtests Similarities and Information as predictors, resulting in the following significant model: $F(2, 258) = 1368.786$, $p < .001$. This model explains 91.3% of the variance in VCI ($\text{adjusted } r^2 = .913$). The standardized regression coefficients for the variables entered into the model were as follows: Similarities $\hat{\beta} = .629$, $p < .001$; Information $\hat{\beta} = .521$, $p < .001$.

2. A similar analysis as in (1) was performed with PRI as criterion variable and the subtests Block Design and Picture Completion as predictors. A significant model emerged: $F(2, 216) = 231.944$, $p < .001$. The model explains 67.9% of the variance in PRI ($\text{adjusted } r^2 = .679$). Both predictors in the model contributed significantly to the outcome: Block Design $\hat{\beta} = .787$, $p < .001$; Picture Completion $\hat{\beta} = .110$, $p = .007$.

3. The validity of PSI was tested in a bivariate regression using the standard method with PSI as the criterion variable and the subtest Coding as the only predictor. The outcome model was significant: $F(1, 258) = 874.428$, $p < .001$, with the predictor Coding explaining 77.2% of the variance in PSI ($r^2 = .772$, $\hat{\beta} = .879$).

As in the full WAIS-IV, the seven-subtest short form WMI consists of two subtests, Digit Span and Arithmetic, which rendered further analyses unnecessary.

Due to the earlier mentioned risk that the correlations emanating from the regression analyses can be misleading if looked at as a typical validity coefficient, a comparison of the full-scale WAIS-IV (FSIQ) and indices versus the values obtained from the prorating method for the seven-subtest SF was conducted using a paired samples $t$-test. Because of the multiple comparisons significance levels were set to $p = .01$ after Bonferroni adjustment. The results revealed no significant difference in FSIQ between these two test batteries. Nevertheless, at index level a significant difference was found between VCI and seven-subtest VCI as well as PRI and seven-subtest PRI (see Tables 3 and 4).

When testing the predictive power of the seven-subtest short form on the two age groups, (1) under 55 years and (2) 55 and over, the following results emerged: group (1) $F(7, 136) = 271.885$, $p < .001$ and 93.0% explained variance in FSIQ (adjusted $r^2 = .930$),
In line with previous reasoning, a paired samples t-test was performed. The results differed between the two age groups. Group (1): no significant difference was found between FSIQ and seven-subtest FSIQ, PSI and seven-subtest PSI and PRI versus seven-subtest PRI. However, a significant difference was found between VCI and seven-subtest VCI.

In the older age group (2): significant differences were found between FSIQ and seven-subtest FSIQ, PRI and seven-subtest PRI, and VCI and seven-subtest VCI. There was no significant difference between PSI and seven-subtest PSI (see Tables 5 and 6).
In order to explore the possibilities of a further reduction of the seven-subtest SF (Ward 1990) hierarchical multiple regression analyses were conducted, with FSIQ as criterion variable. In these analyses the predictors were hierarchically removed, successively eliminating the predictor with the lowest standardized beta coefficient until the breaking point where statistical certainty declined, which occurred at a three-subtest level and below, where there was a reduction of the predictability of FSIQ. The optimal reduction yielded a model consisting of the following predictors: Block Design, Similarities, Arithmetic and Coding revealing a model including one subtest from each index scale (Table 7).

This four-subtest model was significant, $F(4,255) = 401.223$, $p < .001$, and explained 86.1% of the variance in FSIQ (adjusted $r^2 = .861$). All predictors included in the model were significant, see Table 8.

As in the case of the seven-subtest short form, a similar comparison of the complete WAIS-IV FSIQ and indices versus the values obtained from the prorating procedure for the four-subtest short form was conducted using a paired samples $t$-test. All results indicated significant differences except for PSI (Tables 9 and 10).

**Discussion**

**Main results**

The main objective of this study was to investigate whether the seven-subtest short form according to Ward (1990) was statistically valid to use on the Swedish version of the WAIS-IV (Wechsler, 2010). The results of the regression analyses and the paired-samples $t$-test comparing the two short forms’ values with the complete WAIS-IV were somewhat contradictory. According to the regression analyses the seven-subtest short form (Ward, 1990) was a very good predictor of FSIQ in the Swedish adaptation of the WAIS-IV, explaining 93.1% of the variance of FSIQ which is congruent with earlier findings (Meyers et al., 2013; Bulzacka et al., 2016).
Table 7. Standardized regression coefficients for subtests as predictors with FSIQ as criterion variable.

| No. of tests | Block design | Similarities | Coding | Arithmetic | Information | Digit span | Picture completion | Adjusted $r^2$ | $N$  |
|--------------|--------------|--------------|--------|------------|-------------|------------|-------------------|----------------|-----|
| 7            | .310         | .244         | .254   | .238       | .215        | .214       | .041              | .931           | 218 |
| 6            | .319         | .245         | .242   | .251       | .212        | .202       | -                 | .932           | 260 |
| 5            | .323         | .247         | .267   | .326       | .246        | -          | -                 | .903           | 260 |
| 4            | .354         | .307         | .259   | .417       | -           | -          | -                 | .861           | 260 |
Table 8. The unstandardized and standardized regression coefficients for the variables in the suggested four-subtest short form entered into the model predicting FSIQ.

| Variable     | B    | SE B | B   | p   |
|--------------|------|------|-----|-----|
| Block design | 1.489| .122 | .354| <.001|
| Similarities | 1.203| .097 | .307| <.001|
| Arithmetic   | 1.569| .099 | .417| <.001|
| Coding       | 1.117| .108 | .259| <.001|

Table 9. FSIQ and indices test results for WAIS-IV and four-subtest short form.

| Measures  | N  | Mean  | SD   |
|-----------|----|-------|------|
| FSIQ      | 260| 115.670| 11.643|
| fourFSIQ  | 260| 117.540| 13.288|
| VCI       | 261| 118.970| 12.251|
| fourVCI   | 261| 127.540| 17.020|
| PRI       | 261| 113.920| 14.212|
| fourPRI   | 261| 109.720| 16.589|
| PSI       | 260| 108.980| 13.212|
| fourPSI   | 260| 110.010| 14.765|

Table 10. Paired samples t-test between WAIS-IV and four-subtest short form.

| Pair            | Mean | SD   | t-value | df  | p   |
|-----------------|------|------|---------|-----|-----|
| FSIQ - fourFSIQ | -1.865| 5.311| -5.663  | 259 | .000|
| VCI - fourVCI   | -8.567| 9.794| -14.132 | 260 | .000|
| PRI - fourPRI   | 4.192 | 9.386| 7.214   | 260 | .000|
| PSI - fourPSI   | -1.027| 7.059| -2.346  | 259 | .020|

Also, at an index level the validity coefficients of the short form (Ward, 1990) showed strong correlations ranging between adjusted $r^2 = .913$ for VCI and adjusted $r^2 = .679$ for PRI. The results obtained through the prorating method via the paired-samples t-test showed a good fit between the complete WAIS-IV FSIQ and the seven-subtest short form FSIQ supporting the outcome of the regression analysis. Nevertheless, a comparison between the full WAIS-IV indices and the prorated indices of the seven-subtest short form disclosed a lack of consistency. When using the prorating method, the results showed significant differences on VCI as well as PRI and PSI. The PSI and VCI scores obtained from the seven-short form were significantly higher compared to the ones from the full WAIS-IV. For the PRI the situation was opposite, the full WAIS-IV score was significantly higher. This means, contrary to earlier findings in non-Swedish contexts, that the seven-subtest short form is of limited use in predicting the indices in Swedish samples (e. g. Girard et al., 2010; Meyers et al., 2013). An interpretation of these contradictory results emanating from the regression analysis and the paired-samples t-test on an index level is that when reducing
the number of tests that forms separate indices, which in turn creates a main index, the predictive power is challenged. It may be that a good prediction of, in this case FSIQ, is possible, but a consequence of reducing the number of tests forming each individual index is a loss of predictive power at index level. Statistically, regression focuses on the consistency of relations between the measures, not their absolute agreement, which can be one explanation of the nature of the results. In sum, due to this problem clinicians and researchers should use caution when using short forms for profile analyses on index level in a Swedish context.

The second aim was to investigate the applicability of the seven-subtest short form on different age groups. When splitting the sample based on age, group (1) < 55 and group (2) 55 and older, a different pattern occurred. As previously, results from the regression-based analyses showed strong predictive power of the seven-subtest short form versus full WAIS-IV for FSIQ and all the indices in both age groups. The results emanating from the prorating method and associated paired-samples t-test differed between the two age groups. In the younger age group (1), the only significant difference was between VCI and seven-subtest VCI. There were no significant differences in the measures comparing FSIQ ($p = .856$) nor in the indices PRI ($p = .475$) and PSI ($p = .953$). This god fit between the measures indicates that the seven-subtest short form may be suitable when testing a younger sample <55 years in a Swedish context. On the contrary, in the older age group (2) there were significant differences between the measures of FSIQ and also the indices VCI and PCI (FSIQ-sevenFSIQ: $p = .007$, VCI-sevenVCI: $p = .004$ and PCI-sevenPCI: $p < .001$). PSI was the sole index that demonstrated a non-significant difference between full WAIS-IV and the seven-subtest short form. This mismatch in the elderly group (2) cannot be understood by the increase of heterogeneity in cognitive performance levels among the elderly reported in previous studies (Wisdom et al., 2012; Ardila, 2007), because in the present study, no such increase of heterogeneity was observed in the elderly group (2) compared to the younger group (1). One possible interpretation of this lack of increased proliferation of variability in the old age group (2) may be that the sample was split in participants under and over 55 years, rendering the group of elderly persons generally too young, and therefore not being sensitive enough to age related variations. In the present study the number of even older participants was not large enough to create a subsample. This is not to say that the results concerning the old age group (2) of the present study is unreliable, since the group consisted of participants ranging from 55 years to 73 years.

The third objective of this study was to investigate the possibility of a further reduction of the short form (Ward, 1990). The statistically best abbreviation was a four-subtest model including Block Design, Similarities, Arithmetic and Coding, predicting FSIQ with an 87.6% certainty. A particular strength with this model was that it includes one subtest from each index, which ought to be seen as a prerequisite for a global estimate of FSIQ. Girard, et al., (2015) found Coding and Information to be the best dyad to estimate FSIQ and a dyad with Arithmetic also to be highly ranked, in a clinical sample. In another clinical study by Denney et al. (2015) Block Design was consistently identified in several dyads, Similarities included, as the single best predictor of the WAIS-IV FSIQ and the General Ability Index (GAI), with $r^2$ ranging from .79 to .88. Thus, all of the four-subtests here derived from the short form (Ward, 1990) have in previous studies been included in various accurate combinations, which supports the validity and reliability of the proposed four-subtest model to serve as a general measure of WAIS-IV FSIQ. Although the present four-subtest model is not
derived from all ten core tests in the WAIS scale, previous findings indicate that these four subtests may also serve as predictors in an abbreviation based on the full scale (e. g. Girard et al., 2015; Schrimsher, O’Bryant, O’Jile, & Sutker, 2008).

However, contrariwise to the fact that the regression analysis showed that the four-subtest model had a high predictive power with all the predictors contributing significantly to the model, the paired-samples t-test revealed significant differences between the full scale WAIS-IV and four-subtest model (FSIQ-fourFSIQ: p < .001; VCI-fourVCI < .001; PRI-fourPRI: <.001). This obvious discrepancy between these two modes of analyses makes it difficult to generalize the results from the short form level to the full scale level. On one hand the results from the regression analysis is supported by several previous studies (e. g. Girard, et al., 2015; Schrimsher, et al., 2008), on the other hand, the strength of the calculated short forms - full scale correlation may be inflated due to the non-independent measurement error shared by the part (short forms) and its whole (full scale). The obtained correlations can be misleading if looked at as typical validity coefficients (Tellegen & Briggs, 1967). Further, it should be kept in mind that the prorating method to a certain extent implies estimates and not empirical measurements. Having said that, the method has often been used in earlier studies (Abraham, Axelrod & Paolo, 1997; Girard & Christensen, 2008). The contradictory results in this study makes the four-subtest model appear as a questionable instrument. To certify if this contradiction is sample dependent or not an application on another non-clinical sample seems to be necessary.

The development of short forms is not to say that the full scale version of the WAIS should be abandoned. It has long been stressed that short forms may impede pattern analyses (Wechsler, 1944 referred in Olivier, Golden, Acevedo, Sterk, Espinosa, & Spengler, 2013) and should be used with caution in cases of extreme values (Engelhart, Eisenstein, Johnson, & Losonczy, 1999), but are suitable as screening instruments when exact FSIQ-measures are not needed and for research purposes (Sattler & Ryan, 2009; van Ool et al., 2017). It is also important to pay attention to the differences between the group and individual levels. Group results can not automatically may be extrapolated to the individual level (van Ool et al., 2017).

Limitations
The non-clinical sample was not selected according to Swedish census to mirror the population in terms of e.g. age, education, region and gender, which means that the representativeness of the sample is not ascertained. Further, the sample FSIQ was at a high average level, mean 115.58 points, signaling a higher intellectual ability than average according to the normal distribution curve, where the mean is set at 100 IQ points. On the other hand, the index values spanned from borderline level, PSI: 71 IQ-points to a very superior level VCI & WMI: 150 IQ-points indicating a considerable intellectual span in the material, which is characteristic for a normal population (Chevalier et al., 2016). As a whole, the mode of recruitment and the elevated mean FSIQ restrict the generalizability of the results.

Conclusions and suggestion for research
In sum, the results of the study were contradictory. On an FSIQ-level the outcome of the regression analyses and the prorating method were congruent showing that the seven-
subtest model may be an appropriate set in a Swedish context. On index level, there were several cases of mismatches between the results of these two modes of analyses. Due to this condition clinicians and researchers are advised to use this abbreviation with caution for profile analyses on index level. The applicability of the seven-subtest short form on different age groups is questionable. While results being congruent in the younger and middle aged sample (<55 years) suggesting a fit of this short form for this age group in a Swedish setting, the results concerning the elderly group (55 years and older) were incongruent, possibly due to an incapability in capturing the variability in intellectual functioning among the elderly. This suggests a limited applicability of the model for this age group. It is recommended that an analysis focusing on the heterogeneity of cognitive performance among the elderly is made on a sample with higher amounts of participants ≥65 years than in the present study.

Concerning the usefulness of the four-subtest model derived from the seven-subtest abbreviation, the results of the two modes of analyses were incongruent, rendering this model questionable. To certify if this incongruence is sample dependent or not an application on another non-clinical sample seems to be necessary.

Disclosure statement
No potential conflict of interest was reported by the authors.

Ethical agreement
The study was approved by the Regional Ethical Review Board in Stockholm, Dnr: 2017/549-31.

REFERENCES
Abraham, E., Axelrod, B. N., & Paolo, A. M. (1997). Comparison of WAIS-R selected subtest short forms in a mixed clinical population. Assessment, 4, 409–417.
Ardila, A. (2007). Normal aging increases cognitive heterogeneity: Analysis of dispersion in WAIS-III scores across age. Archives of Clinical Neuropsychology, 22(8), 1003–1011. https://doi.org/10.1016/j.acn.2007.08.004
Bowden, S. C., Lissner, D., McCarthy, K. A., Weiss, L. G., & Holdnack, J. A. (2007). Metric and structural equivalence of core cognitive abilities measured with the Wechsler Adult Intelligence Scale-III in the United States and Australia. Journal of Clinical and Experimental Neuropsychology, 29(7), 768–780.
Bowden, S. C., Saklofske, D. H., & Weiss, L. G. (2011a). Augmenting the core battery with supplementary subtests: Wechsler Adult Intelligence Scale—IV measurement invariance across the United States and Canada. Assessment, 18, 133–140.
Bowden, S. C., Saklofske, D. H., & Weiss, L. G. (2011b). Invariance of the measurement model underlying the Wechsler Adult Intelligence Scale-IV in the United States and Canada. Educational and Psychological Measurement, 71(1), 186–199.
Bulzacka, E., Meyers, J. E., Boyer, L., Le Gloahec, T., Fond, G., Szöke, A., … & Schürhoff, F. (2016). WAIS-IV seven-subtest short form: Validity and clinical use in Schizophrenia. Archives of Clinical Neuropsychology, 31:915–925.
Chevalier, T. M., Stewart, G., Nelson, M., McInerney, R. J., & Brodie, N. (2016). Impaired or not impaired, that is the question: Navigating the challenges associated with using canadian normative data in a comprehensive test battery that contains American tests. Archives of Clinical Neuropsychology, 31, 446–455. doi: 10.1093/arclin/acw031.
Crawford, J. R., Allum, S., & Kinion, J. E. (2008). An index-based short form of the WAIS-III with accompanying analysis of reliability and abnormality of differences. *British Journal of Clinical Psychology, 47*, 215–237.

Daseking, M., Petermann, F., & Waldmann, H. C. (2014). Schätzung der allgemeinen Intelligenz mit einer Kurzform der WAIS-IV bei neurologischen Fragestellungen. *Aktuelle Neurologie, 41*(06), 349–355.

Denney, D. A., Ringe, W. K., & Lacritz, L. H. (2015). Dyadic short forms of the Wechsler Adult Intelligence Scale-IV. *Archives of Clinical Neuropsychology, 30*, 404–412. doi: 10.1177/1073191114551551

Engelhart, C. I., Eisenstein, N., Johnson, V., & Losonczy, M. (1999). Comparison of linear equating and prorated short forms for estimating WAIS-R FSIQ in a neuropsychological population. *The Clinical Neuropsychologist, 13*, 95–99.

Escobedo, P. S., & Hollingworth, L. (2009). Annotations on the use of the Mexican norms for the WAIS-III. *Applied neuropsychology, 16*, 223–227.

Girard, T. A., Axelrod, B. N., Patel, R., & Crawford, J. R. (2015). Wechsler Adult Intelligence Scale–IV Dyads for estimating global intelligence. *Assessment, 22*, 441–448. doi: 10.1177/1073191114551551

Girard, T. A., Axelrod, B. N., & Wilkins, L. K. (2010). Comparison of WAIS-III short forms for measuring index and full-scale scores. *Assessment, 17*(3), 400–405.

Girard, T. A., & Christensen, B. K. (2008). Clarifying problems and offering solutions for correlated error when assessing the validity of selected-subtest short forms. *Psychological Assessment, 20*, 76–80.

Meyers, J. E., Zellinger, M. M., Kockler, T., Wagner, M., & Miller, R. M. (2013). A validated seven-subtest short form for the WAIS-IV. *Applied Neuropsychology: Adult, 20*, 249–256. doi:10.1080/09084282.2012.710180

Olivier, T. W., Golden, C. J., Acevedo, A., Sterk, V. I., Espinosa, K. M., & Spengler, K. M. (2013). WAIS-IV index and full scale intelligence quotient score differences between standard and prorated scoring and arched procedures. *Archives of Assessment Psychology, 3*, 57–77.

Paolo, A. M., & Ryan, J. J. (1993). WAIS-R abbreviated forms in elderly: A comparison of the Satz-Mogel with a seven-subtest short form. *Psychological Assessment, 5*, 425–429.

Pilgrim, B. M., Meyers, J. E., Bayless, J., & Whetstone, M. M. (1999). Validity of the Ward seven-subtest WAIS-III short form in a neuropsychological population. *Applied Neuropsychology, 6*, 243–246.

Roivainen, E. (2013). Are cross-national differences in IQ pro...? A comparison of Finnish and US WAIS norms. *International Journal of Testing, 13*, 140–151. http://dx.doi.org/10.1080/15305058.2012.657315

Ryan, J. J., Kreiner, D. S., Gontkovsky, S. T., & Glass Umfleet, L. (2015). Classification accuracy of sequentially administered WAIS-IV short forms. *Applied Neuropsychology: Adult, 22*, 409–414.

Sattler, J. M., & Ryan, J. J. (2009). *Assessment with the WAIS-IV*. San Diego: Jerome M. Sattler Publisher Inc.

Schrimsher, G. W., O’Bryant, S. E., O’Jile, J. R., & Sutker, P. B. (2008). Comparison of tetradic WAIS-III short forms in predicting full scale IQ scores in neuropsychiatric clinic settings. *Journal of Psychopathology and Behavioral Assessment, 30*(3), 235–240.

Silverstein, A. B. (1990). Short forms of individual intelligence tests. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2*, 3–11.

Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. Washington, D.C.: American Chemical Society.

Tellegen, A., & Briggs, P. F. (1967). Old wine in new skins: Grouping Wechsler subtests into new scales. *Journal of Consulting Psychology, 31*(5), 499–506.

Úbeda, R., Fuentes, I., & Dasi, C. (2016). Revisión de las formas abreviadas de la Escala de Inteligencia de Wechsler para Adultos. *Psychology, Society, & Education, 8*, 81–92.

van Ool, J. S., Hurks, P. P., Snoeijen-Schouwenaars, F. M., Tan, I. Y., Schelhaas, H. J., Klinkenberg, S., … & Hendriksen, J. G. (2017). Accuracy of WISC-III and WAIS-IV short forms in patients with neurological disorders. *Developmental Neurorehabilitation, 21*(1), 1–7. doi:10.1080/17518423.2016.1277799

Ward, L. C. (1990). Prediction of verbal, performance, and full scale IQs from seven subtests of the WAIS-R. *Journal of Clinical Psychology, 46*, 436–440.

Ward, L. C., Bergman, M. A., & Hebert, K. R. (2012). WAIS-IV subtest covariance structure: Conceptual and statistical considerations. *Psychological Assessment, 24*(2), 328.
Wechsler, D. (1989). *Wechsler preschool and primary scale of intelligence–revised*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1991). *Wechsler intelligence scale for children–third edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale–third edition*. San Antonio, TX: NCS Pearson.

Wechsler, D. (2002). *Wechsler preschool and primary scale of intelligence–third edition*. San Antonio, TX: Harcourt Assessment.

Wechsler, D. (2003). *Wechsler intelligence scale for children–fourth edition*. San Antonio, TX: Harcourt Assessment.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–fourth edition (WAIS–IV)*. San Antonio, TX: NCS Pearson.

Wechsler, D. (2010). *WAIS IV: Wechsler Adult Intelligence Scale–Fourth edition. Swedish version*. Toronto, Ontario: NCS Pearson Inc.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WAIS-IV and clinical validation of the four-and five-factor interpretative approaches. *Journal of Psychoeducational Assessment*, 31, 94–113. doi: 10.1177/0734282913478030

Wisdom, N. M., Mignogna, J., & Collins, R. L. (2012). Variability in Wechsler Adult Intelligence Scale-IV subtest performance across age. *Archives of Clinical Neuropsychology*, 27, 389–397. doi: 10.1093/arclin/ac041