Available online at www.sciencedirect.com
ScienceDirect
Intelligence 36 (2008) 48–67
INTELLIGENCE

Sex differences on the WISC-R in Belgium and The Netherlands

Sophie van der Sluis [a,*], Catherine Derom [b], Evert Thiery [c], Meike Bartels [a], Tinca J.C. Polderman [a,d], F.C. Verhulst [d], Nele Jacobs [c], Sofie van Gestel [c], Eco J.C. de Geus [a], Conor V. Dolan [e], Dorret I. Boomsma [a], Danielle Posthuma [a]

[a] *Department of Biological Psychology, VU University Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands*
[b] *Department of Human Genetics, University Hospital Gasthuisberg, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium*
[c] *Association for Scientific Research in Multiple Births, B-9070 Destelbergen, Belgium*
[d] *Department of Child and Adolescent Psychiatry, Erasmus MC- Sophia, Dr. Molewaterplein 60, 3015 GJ, Rotterdam, The Netherlands*
[e] *Department of Psychology, FMG, University of Amsterdam, Roeterstraat 15, 1018 WB, Amsterdam, The Netherlands*

Received 6 May 2006; received in revised form 15 January 2007; accepted 17 January 2007
Available online 22 February 2007

Abstract

Sex differences on the Dutch WISC-R were examined in Dutch children (350 boys, 387 girls, age 11–13 years) and Belgian children (370 boys, 391 girls, age 9.5–13 years). Multi-group covariance and means structure analysis was used to establish whether the WISC-R was measurement invariant across sex, and whether sex differences on the level of the subtests were indicative of sex differences in general intelligence (*g*). In both samples, girls outperformed boys on the subtest Coding, while boys outperformed girls on the subtests Information and Arithmetic. The sex differences in the means of these three subtests could not be accounted for by the first-order factors Verbal, Performance, and Memory. Measurement invariance with respect to sex was however established for the remaining 9 subtest. Based on these subtests, no significant sex differences were observed in the means of the first-order factors, or the second-order *g*-factor. In conclusion, the cognitive differences between boys and girls concern subtest-specific abilities, and these sizeable differences are not attributable to differences in first-order factors, or the second-order factor *g*.

Keywords: Intelligence; Sex-differences; Multi-group covariance and mean structure analysis; Measurement invariance

1. Introduction

Sex differences on the WISC-R have been studied in the WISC-R standardization samples of the USA, Scotland, The Netherlands, and China, and in data from Mauritius, New Zealand, and Belgium (e.g., Born & Lynn, 1994; Dai & Lynn, 1994; Grégoire, 2000; Jensen & Reynolds, 1983; Lynn & Mulhern, 1991; Lynn, Riane, Venables, Mednick, & Irwing, 2005). The results are largely comparable across countries. Consistently, large differences favoring girls are reported regarding the subtest Coding (effect sizes about .5), and large differences favoring boys are reported regarding the subtest Information (effect sizes about .35). In addition, girls sometimes outperform boys on the subtest Digit Span, but these differences are usually small and statistically insignificant. Boys score slightly higher than girls on all other subtests, and even though these differences are sometimes statistically significant,

* Corresponding author.
E-mail address: s.van.der.sluis@psy.vu.nl (S. van der Sluis).

0160-2896/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2007.01.003

the differences are often small, with effect sizes ranging between .00 and .20.

In all these studies, WISC-R subtest scores and factor scores have been compared directly between boys and girls. Yet it has never been established whether the factor structure of the WISC-R is actually comparable or 'measurement invariant' across sex (see below). The interpretation of group differences in subtest- or factors scores may be complicated greatly if the underlying factor structure differs between the groups. That is, if a test battery does not measure the same construct(s) in different groups, then group differences in test scores representing first or higher order factors are difficult to interpret. The aim of the present study is to find out whether the WISC-R is measurement invariant across sex in children before comparing subtest and factor scores between boys and girls.

The factor structure underlying the WISC-R has been studied in clinical and non-clinical samples (e.g., Anderson & Dixon, 1995; Burton et al., 2001; Donders, 1993; Huberty, 1987; Kush et al., 2001; Meesters, van Gastel, Ghys, & Merckelbach, 1998; Wright & Dappen, 1982). Principal component analyses (PCA, e.g., Born & Lynn, 1994; Lynn & Mulhern, 1991; Rushton & Jensen, 2003), exploratory factor analyses (EFA, e.g., Dolan, 2000; Dolan & Hamaker, 2001; Kush et al., 2001), and confirmatory factor analyses (CFA, e.g., Burton et al., 2001; Dolan, 2000; Dolan & Hamaker, 2001; Keith, 1997; Kush et al., 2001; Oh, Glutting, Watkins, Youngstrom, & McDermott, 2004) have yielded either a two factor ('Verbal' and 'Performance'), or a three factor solution ('Verbal', 'Performance', and 'Memory', also known as 'Freedom from distractibility'). In these models, general intelligence ('*g*') was either operationalized as the first principal component (PCA), or as a second-order factor (CFA).

Given the assumption that these latent factors underlie the performance on the level of the subtests, one question of interest is whether the observed sex differences at the level of the subtests are a function of differences in *g*, or of differences on the level of the broad primary factors of intelligence (e.g., Verbal intelligence, Performance intelligence and Memory). However, it may also be the case that the subtest differences are not attributable to common factor differences, but rather are a manifestation of differences in the specific ability that the subtest taps.

If boys and girls differ with respect to the mean on a given subtest, and this difference cannot be explained by the mean differences on the latent factor, which is supposed to underlie performance on the subtest, then the subtest may be viewed as biased with respect to sex. The term bias does not imply that the observed mean difference is not real, rather the term, as used here, implies that the mean difference on the subtest is greater or smaller than that expected on the basis of the latent factor mean difference. According to this definition, the term bias refers to the subtest as an indicator of the common factor, which the subtest is supposed to measure. For example, it has been established that the Information subtest of the WAIS is biased with respect to sex. Specifically, the male advantage on this subtest, which is supposed to measure general knowledge, is too large to be accounted for by the common factor Verbal Comprehension (e.g., Dolan et al., 2006; Van der Sluis et al., 2006). The difference is not indicative of a difference with respect to Verbal Comprehension. However, it may well be indicative of a true male advantage in general knowledge.

Establishing the exact nature of an observed (subtest) mean difference is important in the light of theories, in which sex differences are attributed to latent mean differences (e.g., a difference in Verbal Comprehension, or a difference in *g*). In previous studies aimed at identifying the source(s) of the sex differences, PCA was mostly used to investigate sex differences on the factors underlying intelligence. Sex differences were evaluated by calculating weighted linear combination of the subtests means, where the subtests' factor loadings served as weights (e.g., Born & Lynn, 1994; Jensen & Reynolds, 1983; Lynn, Fergusson, & Horwood, 2005; Lynn & Mulhern, 1991; Lynn, Riane, et al., 2005). The general finding of these studies is that boys score higher on the Verbal and Performance factors, while girls score higher on the Memory factor. With respect to general intelligence, operationalized as the first principal component, boys usually score higher than girls, but effect sizes are often small (about .10), and the difference is not always statistically significant. When expressed on the conventional IQ-scale with a mean of 100 and standard deviation of 15, these sex differences range from 1 to 6 IQ points (e.g., Lynn, Fergusson, et al., 2005; Lynn, Riane, et al., 2005). All these results are however based on samples with a broad age-range (6–16 years), and it remains to be seen whether the factor structure of the WISC-R, and the effects reported for the (factor) means, are stable across age.

One obvious problem concerning this PCA-based method of studying sex differences is that sex differences on the level of the weighted means of the observed subtest scores may be due to one or just a few of many subtests. For example, boys may outperform girls on the Verbal factor only because they outperformed girls on the subtest Information, while their performance on the other verbal subtests may even be inferior. In that case, it