

KRYSTYNA SOCHACKA

State Higher Vocational School in Nowy Sącz

Institute of Pedagogy

A COMPARISON OF WISC-R AND WAIS-R (PL) SCORES OF CHILDREN AND ADOLESCENTS IN A LONGITUDINAL STUDY

D. Wechsler's tests were used to examine the same group of subjects within the intellectual norm ($N = 31$) twice over an interval of 13 years: the *Wechsler Intelligence Scale for Children* (WISC-R; the subjects' mean age was 8 years) and the *Wechsler Adult Intelligence Scale* (WAIS-R; mean age – 21 years). Each measurement was carried out by the same person. Based on 62 protocols (answer sheets), intelligence quotients, scaled subtest scores, and factor scores obtained using WISC-R were compared both with the normalized version of WAIS-R (PL) (2004) and with the version of WAIS-R (PL) from before normalization (1996). Statistically significant differences were found between the results obtained using WISC-R and the normalized version of WAIS-R (PL). There is no such difference between WISC-R and the earlier version of WAIS-R (PL), from before normalization. Implications for clinical practice are discussed.

Keywords: IQ stability, intelligence tests, WISC-R, WAIS-R (PL).

David Wechsler's scales rank among the most important intelligence tests. The following are available in Poland: the 1974 *Wechsler Intelligence Scale for Children-Revised* (WISC-R; adapted by Matczak, Piotrowska, & Ciarkowska, 1991, 1997, 2008), as well as the 1981 *Wechsler Adult Intelligence Scale* –

Address for correspondence: KRYSTYNA SOCHACKA – Institute of Pedagogy, State Higher Vocational School in Nowy Sącz, ul. Chruślicka 6, 33-300 Nowy Sącz; e-mail: k.sochacka@uwb.edu.pl

In my study I have used the results of the examination of children using WISC-R, carried out as part of research project no. H01F 053 14 financed by the Polish State Committee for Scientific Research (KBN) in 1998-2000.

I would like to thank Ms. Elżbieta Sanocka, M.A. – a psychologist at the Psychological and Pedagogical Counseling Center in Nysa – for inspiration and for consultation concerning the content of this article.

Revised (WAIS-R; adapted by Brzeziński, Gaul, Hornowska, Machowski, & Zakrzewska, 1996) and its renormalized version (Brzeziński et al., 2007), in which, among other modifications, the procedure of computing scaled scores was changed¹ (Zakrzewska, 2001). These are not new versions of Wechsler's scales, and they are no longer used in most countries now. When it comes to the advantages of their Polish adaptations, it is stressed that they meet the conditions of reliability, validity, and normalization (Krasowicz-Kupis & Wiejak, 2006).

What has become an object of interest is the comparability of scores obtained in tests for children and adults. The acknowledgment that the level of intellectual competence constitutes a relatively stable human characteristic that distinguishes a person from others (Nęcka, 2003) gives rise to the expectation that intelligence test scores in different periods of life will be similar. Although it is subject to developmental, degenerative, or civilizational changes, intelligence should maintain a stable position in consecutive measurements against the background of the age group. An intelligence quotient (IQ) of 100 is equal to the average in a given population, and in that sense it is a constant value. Longitudinal studies confirmed the stability of test scores over the periods of elementary, secondary, and higher education (Anastasi & Urbina, 1999). Studies of population groups (Flynn, 2006) showed that, as the norms become older, there is an increase in test scores (the Flynn effect). For the Wechsler scales, the rate of this increase is about 0.3 points a year. The most overestimated IQ values were observed in people with a low level of intelligence.

American researchers have demonstrated on numerous occasions that WISC-R and WAIS-R scores are not equivalent (Spitz, 1989)². The scores were higher in measurements using the scale for adults (WAIS-R) than in those using the scale for children (WISC-R), most often on the *Verbal Scale* and on the *Full Scale* (Grace & Sweeney, 1986; Vance, Brown, Hankins, & Furgerson, 1987; Wilcoxon, 1982). On the other hand, the equivalence of the tests was confirmed by Sattler, Polifka, Polifka, and Hilsen (1984). These authors found no significant differences between two tests completed by adolescents over an interval of four years. Higher scores were also obtained in WISC-R than in WAIS-R (Slate, Frost, & Cross, 1990), which was explained as stemming from WISC-R norms being older compared to WAIS-R norms.

¹ Tables for the conversion of raw scores into scaled scores were developed for each age group – not, as previously, in relation to the scores of the reference group composed of people aged 20-34 years with the highest intellectual level.

² The references are made to the literature of the 1980s, when these scales were in use before they were superseded in the 1990s by newer versions (WISC-III and WAIS-III).

Sixteen-year-olds were tested the most often, since both scales contain norms for this age. Contrary to the expectations that the tasks from the test for children would be easier, 16-year-olds scored higher in the test for adults. The differences depended on the level of intelligence and increased at low IQ values. Wechsler (1981) demonstrated that in a broad range of average scores WISC-R and WAIS-R for 16-year-olds remain comparable, while discrepancies are observed in the case of examinees with a low intellectual level, who score higher in WAIS-R, and in examinees with the highest intelligence, who score higher in WISC-R.

A discrepancy between the two scales is also observed by Polish practicing psychologists. According to Aleksandra Jaworowska (2011), it stems from the different conventions of computing the scores. A participant aged 16.5 years who solves one task from each test in WISC-R will score 10 scaled points and an IQ below 40, whereas in WAIS-R he or she will score 17 points and an IQ of 46. This is a serious assessment problem particularly in a situation when students diagnosed for many years using WISC-R as having a moderate intellectual disability are re-examined using WAIS-R (PL) after finishing school or coming of age and obtain scores indicating a light degree of disability even though the level of their functioning at school or in society has not changed. Such a psychometric diagnosis may have serious administrative consequences; for instance, it may deprive a person of eligibility for social pension. It is worth noting that this problem appeared with the 2004 renormalized version of WAIS-R (PL). Practice shows, however, that in a problematic situation of a large discrepancy between WISC-R and WAIS-R (PL) scores it is enough to return to the 1996 version of WAIS-R (PL) to obtain a level of cognitive ability for a particular person that is comparable to that measured using WISC-R.

The assessment problems of practitioners, the diverse opinions of American researchers, and the lack of Polish studies concerning the stability of test achievements in childhood and adolescence have all contributed to my decision to address this issue in research. The discrepancy between the scales in intellectually disabled people has been confirmed by American researchers. It is also signaled by practicing psychologists. When it comes to scores within the norm, reports are ambiguous. Therefore, I sought the answer to the question of whether in the average score band WISC-R and WAIS-R (PL) will be correspond, or whether, as in the case of low scores, they will differ significantly.

Following the suggestions of practicing psychologists concerning the differences between the 1996 and 2004 versions of WAIS-R (PL), I formulated the questions and hypotheses separately for each version of the test. Taking into account the cited findings reported by Wechsler (1981) and others (cf. Anastasi

& Urbina, 1999) and considering that practitioners' reports relate to low scores, I assumed that in the group of people within the intellectual norm there would be no differences between WISC-R and the 2004 version of WAIS-R (PL). Likewise, I assumed that in this group there would be no significant differences between WISC-R and the 1996 version of WAIS-R (PL), either. I tested the above hypotheses in a longitudinal study, comparing WISC-R scores with those obtained in WAIS-R (PL), both the renormalized version (2004) and the one before renormalization (1996). I devoted special attention to the implications for psychological practice.

METHOD

The same people ($N = 31$, 20 women and 11 men) were examined twice, using WISC-R at the age of about 8 years and using WAIS-R (PL) at about 21 years. Table 1 illustrates the structure of the sample in the second examination.

Table 1
Structure of the Sample

Studies		Women	Men	Total	
Nonstudent		3	1	4	
Nonpublic university	Part-time	1	1	2	6
	Full-time	1	3	4	
Public university	Part-time	4	2	6	21
	Full-time	11	4	15	
Total		20	11	31	

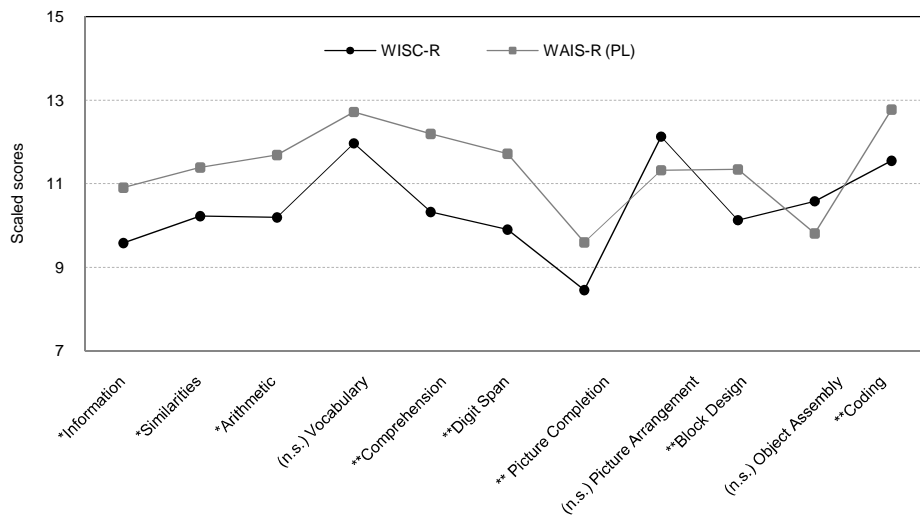
I computed test results for adults in two ways: (1) in accordance with the 2004 instruction and norms for WAIS-R (PL) and (2) in accordance with the 1996 instruction and norms for WAIS-R (PL). This was possible due to small changes in the research procedure between the two versions and fundamental changes in the way of scaling the scores. I compared the intelligence quotients obtained in the *Full*, *Verbal*, and *Nonverbal (Performance) Scales*, scaled scores in each test, and factor scores. The values for factors were determined in accordance with the three-factor model (Zakrzewska, 2000). A comparison of mean values made it possible to check to what extent a given trait remained stable in time. Correlation coefficients showed to what extent the participants retained

their position against the background of the group in terms of the level of the investigated characteristic. In the analysis of data, I used the SPSS statistical package. The significance of differences between WISC-R and WAIS-R (PL) scores was assessed using Student's *t*-test (*t*) for dependent samples, and the covariance of results was illustrated using Pearson's correlation coefficient (*r*).

RESULTS

WISC-R and the 2004 renormalized version of WAIS-R (PL)

The comparison of mean IQ values revealed significant differences in favor of WAIS-R (PL) in the Full Scale and in the Verbal Scale. In the case of the Performance Scale, the difference is not statistically significant. The correlation coefficients turned out to be significant. The positive relationship between the scores in the two tests means that, in terms of the investigated characteristic, the participants retain their position against the background of the group. Results in each test are presented in Figure 1.



Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure 1. Profiles of WISC-R and the 2004 version of WAIS-R (PL).

In most tests, the mean values are significantly higher in the case of WAIS-R (PL). The scores in the two scales are similar in three tests (*Vocabulary*, *Picture Arrangement*, *Object Assembly*). The course of the curves illustrating the measurements at two points in time is similar and changes occur unidirectionally. This covariance is significant in the case of the following subtests: *Information* ($r = .46, p < .01$), *Similarities* ($r = .59, p < .01$), *Block Design* ($r = .58, p < .01$), *Coding* ($r = .56, p < .01$), *Picture Arrangement* ($r = .38, p < .05$), and *Object Assembly* ($r = .44, p < .05$). Pairs of factor scores on *Verbal Comprehension* ($t = -2.95, p < .01, r = .79, p < .001$) as well as *Memory and Freedom from Distractibility* ($t = -4.09, p < .001, r = .47, p < .01$) differ significantly in favor of WAIS-R (PL). The difference between the mean scores on *Perceptual Organization* ($t = -.52$ ns., $r = .60, p < .001$) is not significant. The correlations are significant, and the relationship between each pair of factor scores is positive.

WISC-R and the 1996 version of WAIS-R (PL)

The procedure of converting raw scores into scaled scores in the 1996 version of WAIS-R (PL) caused a considerable decrease in scores compared to the renormalized version in the *Full Scale* ($t = 13.25, p < .001$), the *Verbal Scale* ($t = 14.42, p < .001$), and the *Performance Scale* ($t = 6.78, p < .001$). As a result, the differences between *Full-Scale* and the *Verbal IQs* decreased considerably, becoming statistically nonsignificant. The decrease in *Performance Scale* scores resulted in the previously nonsignificant differences growing, in favor of WISC-R this time. In three tests of the *Performance Scale* (*Picture Completion*, *Block Design*, and *Coding*) the differences remained in favor of WAIS-R (PL). As previously, the correlation coefficients are statistically significant and indicate a positive relationship between the tests. Figure 2 illustrates specific tests of the *Verbal* and *Nonverbal* Scales.

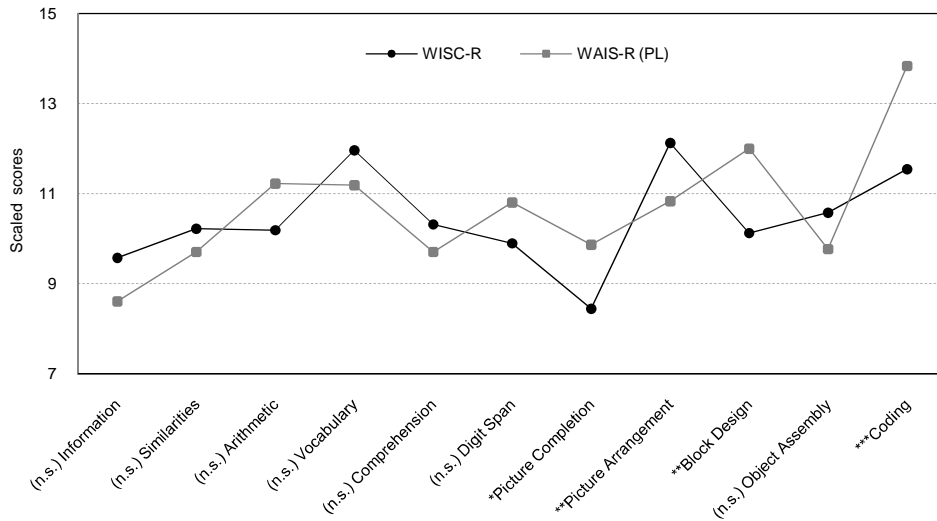


Figure 2. Profiles of WISC-R and the 1996 version of WAIS-R (PL).

Differences between the *Verbal Scale* tests of WISC-R and the 1996 version of WAIS-R (PL) are not statistically significant, while those between *Performance Scale* tests vary. They differ significantly in favor of the test for children in the case of *Picture Arrangement* and in favor of the test for adults in the case of three subtests: *Picture Completion*, *Block Design*, and *Coding*. Correlation coefficients are significant in: *Similarities* ($r = .63, p < .001$), *Information* ($r = .52, p < .01$), *Digit Span* ($r = .55, p < .001$), *Block Design* ($r = .59, p < .01$), *Coding* ($r = .56, p < .01$), *Object Assembly* ($r = .51, p < .01$), and *Picture Arrangement* ($r = .43, p < .05$). The difference between pairs of factor scores in *Verbal Comprehension* ($t = 3.46, p < .01, r = .70, p < .001$) is in favor of WISC-R, which is the opposite of the situation in the 2004 version of WAIS-R (PL) and different than in the comparison of *Verbal IQs*. The difference between the *Perceptual Organization* factor scores ($t = 0.47$ ns., $r = .42, p < .05$) is nonsignificant, just like in the comparison with the more recent version of WAIS-R (PL), but also differently than in the comparison of *Performance IQs*. The third factor, *Memory and Freedom from Distractibility* ($t = -3.67, p < .01, r = .52, p < .01$), comes out in favor of adolescents in the case of both ways of converting scores, which makes it legitimate to acknowledge real progress in this respect between the compared examinations.

DISCUSSION

The comparison of scores obtained by individuals within the intellectual norm, tested using WISC-R and WAIS-R (PL), the version currently used in Poland, suggests that during the time between the measurements the group developed their verbal abilities, while the nonverbal abilities remained at the same level. This result is consistent with the findings of American researchers and contradicts the supposition that the two scales are equivalent when it comes to average scores. The comparison of the scale for children and the 1996 version of the scale for adults yielded different results. They suggest that the performance abilities of the group decreased between the measurements, while verbal abilities remained at a comparable level.

The differences between the scores of children and young people, determined by the version of WAIS-R (PL) that was administered, cannot be explained by the Flynn effect. According to Flynn (2006), older norms inflate scores, whereas the study has revealed the reverse: namely, that the scores compared to the older norms for the 1996 version of WAIS-R (PL) were significantly lower than those compared to the more recent norms for the 2004 version of WAIS-R (PL). It thus turned out that not only in the low IQ range but also in the average IQ range the scores in the 2004 version of WAIS-R (PL) were significantly higher compared to WISC-R. Observations of this kind have not been reported by practitioners. In the average score range, such a change may escape the attention of assessment psychologists. In individuals within the intellectual norm, progress in mental development is much less often monitored using tests. Moreover, in interpretation it is easier to assume that changes are a natural consequence of a particular person's development.

Uncertainty arises, however, as to whether the increase in WAIS-R (PL) scores can be treated as an effect of changes in the level of cognitive functioning between childhood and adolescence. A positive answer is supported, on the one hand, by the dynamics of development in the period investigated, and on the other by the strict association of test scores with the abilities developed in schools. Measurements usually show an increase in scores in people who continue education for a longer time (Anastasi & Urbina, 1999). The participants in the study were students, who practiced these academic skills between the measurements. It is stated in the test manual (Brzeziński et al., 2007) that the variable differentiating WAIS-R (PL) scores the most strongly is education level. However, assuming the constancy of IQ during the life span, changes observed in

the repeated test examination can also be attributed to the weakness of the measurement instruments.

The concept of IQ constancy relates to populations, not to individuals. Still, the question arises whether the practice of using the older, 1996 version of WAIS-R (PL) in a situation of a problematic discrepancy between the 2004 version of WAIS-R (PL) and WISC-R can be a justified practice. According to Anastasi and Urbina (1999), as long as the measure serves to formulate hypotheses, its use may be justified. In individual assessment, a psychologist has various kinds of data at his or her disposal – for example, data from observation or interview, which he or she combines with test scores to obtain a comprehensive picture of the examinee. This prevents excessive generalization of test scores and partly explains the insistence on using tests regardless of their psychometric properties. Interaction during examination provides the experienced assessment psychologist with considerably more information about the examinee's strong and weak points than IQ alone.

Relying on the psychometric criterion only can lead to simplifications or even to erroneous conclusions, as numerous authors point out. These authors acknowledge that tests are helpful in making a diagnosis, particularly at the preliminary stage, but it is full psychological and clinical examination that should be conclusive. The responsibility for the accurate use of a test rests with the person interpreting it, and the interpretation carries with it the responsibility for the consequences (Messick, 2005; *Standardy dla testów*, 2007). Anastasi and Urbina (1999) believe that to decide only on the basis of tests is to abuse them. Tests constitute just one source of data; they are not conclusive instruments, and decisions ought to be made by people.

Making Wechsler scales for measuring intelligence available to Polish psychologists is of invaluable importance to psychological assessment. The scales have been localized to the Polish cultural realities but they have retained the "spirit" of the original (Brzeziński et al., 1996, 2007) and, together with it, the problems signaled both by American researchers and by Polish psychologists. More recent and better versions of the tests will probably be free of these problems.

REFERENCES

- Anastasi, A., & Urbina, S. (1999). *Testy psychologiczne*. Warsaw, PL: Psychological Test Laboratory of the Polish Psychological Association.

- Brzeziński, J., Gaul M., Hornowska, E., Jaworowska, A., Machowski, A., & Zakrzewska, M. (2007). *Skala Inteligencji D. Wechslera dla dorosłych. Wersja zrewidowana – Renormalizacja, WAIS-R (PL)*. Warsaw, PL: Psychological Test Laboratory of the Polish Psychological Association.
- Brzeziński, J., Gaul, M., Hornowska, E., Machowski, A., & Zakrzewska, M. (1996). *Skala Inteligencji D. Wechslera dla dorosłych. Wersja zrewidowana WAIS-R (PL)*. Warsaw, PL: Psychological Test Laboratory of the Polish Psychological Association.
- Flynn J. (2006). Tethering the elephant. Capital cases, IQ, and the Flynn effect. *Psychology, Public Policy, and Law*, 12, 170-189.
- Grace W., & Sweeney M. (1986). Comparisons of the $P > V$ sign on the WISC-R and WAIS-R in delinquent males. *Journal of Clinical Psychology*, 42, 173-176.
- Jaworowska, A. (2011). Starzenie się norm w testach inteligencji. Efekt Flynna na przykładzie wyników WISC-R. In K. Wiejak & G. Krasowicz-Kupis (Eds.), *Kliniczne zastosowania skal inteligencji D. Wechslera* (pp. 17-31). Warsaw, PL: Psychological Test Laboratory of the Polish Psychological Association.
- Krasowicz-Kupis, G., & Wiejak, K. (2006). *Skala Inteligencji Wechslera dla dzieci (WISC-R) w praktyce psychologicznej*. Warsaw, PL: Wydawnictwo Naukowe PWN.
- Matczak, A., Piotrowska, A., & Ciarkowska, W. (1991, 1997, 2008). *Skala Inteligencji D. Wechslera dla dzieci – Wersja zmodyfikowana (WISC-R)*. Warsaw, PL: Psychological Test Laboratory of the Polish Psychological Association.
- Messick, S. (2005). Trafność testu a etyka oceny (diagnozy). In J. Brzeziński (Ed.), *Trafność i rzetelność testów psychologicznych* (pp. 460-485). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Nęcka, E. (2003). *Inteligencja. Geneza – Struktura – Funkcje*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Sattler, J. M., Polifka, J., Polifka, S., & Hilsen, D. (1984). A longitudinal study of the WISC-R and WAIS-R with special education students. *Psychology in the Schools*, 3, 294-295.
- Slate, J. R., Frost, J., & Cross, B. (1990). Comparability of WISC-R and WAIS-R scores for a sample of college students with learning disabilities. *Learning Disability Quarterly*, 13, 205-208.
- Spitz, H. (1989). Variations in Wechsler Interscale IQ. Disparities at different levels of IQ. *Intelligence*, 13, 157-167.
- Standardy dla testów stosowanych w psychologii i pedagogice* (2007). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Vance, H., Brown, W., Hankins, N., & Furgerson, S. (1987). A comparison of the WISC-R and the WAIS-R with special education students. *Journal of Clinical Psychology*, 43, 377-380.
- Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale – Revised*. New York: Psychological Corporation.
- Wilcoxon, K. (1982). *A comparison of the WAIS-R and WISC-R for sixteen-year-old learning disabled and educable mentally handicapped students*. A thesis presented to the Department of Psychology, Emporia State University. Retrieved from <https://esirc.emporia.edu/handle/123456789/2120>
- Zakrzewska, M. (2000). Trzy różne ilorazy inteligencji: interpretacja polskiej wersji Skali Inteligencji dla Dorosłych [WAIS-R (PL)]. *Czasopismo Psychologiczne*, 6, 159-169.
- Zakrzewska, M. (2001). Poziom wykonania testów WAIS-R (PL) w różnych grupach wiekowych. *Roczniki Psychologiczne*, 4, 257-279.