A R T I C L E S

ANNALS OF PSYCHOLOGY/ROCZNIKI PSYCHOLOGICZNE
2024, XXVII, 2, 103–118
DOI: https://doi.org/10.18290/rpsych2024.0006

ITEM ANALYSIS OF SOCIOCULTURAL INFLUENCES IN VERBAL ABILITY TESTING USING THE WOODCOCK-JOHNSON IV COG

Tomáš Mrhálek[1], Alena Hricová[2], and Stanislav Ondrášek[2]

[1] Department of Pedagogy and Psychology, University of South Bohemia
[2] Faculty of Health and Social Studies, University of South Bohemia

The article presents the results of a study that examines the impact of sociocultural environment on the development of verbal skills in Romani children aged 7–11 years, with a particular emphasis on children from socially excluded localities. The results are based on a verbal test derived from the Woodcock-Johnson Cognitive Ability Test (WJ IV COG), which was processed using item analysis according to the Rasch model. The main results are based on analysing the differential function of individual items within the verbal tests within the research group ($N = 399$) compared to the normative sample of the Czech population ($N = 936$). A secondary goal is to compare Romani children originating from socially excluded areas ($N = 204$) with those from non-excluded areas ($N = 195$). Regarding item comparisons between the majority and sample groups, they reveal significant disparities in 14 items (70%) of the Synonyms subtest and 9 items (45%) of the Antonyms subtest. The results highlight the crucial role of a more limited vocabulary, particularly in formal linguistic expressions by Romani children. The presented words representing differently functioning items can also serve as a basis for linguistic analysis and can be applied in assessing educational needs. When comparing the group of children based on their residence in a socially excluded locality, significant differences in the DIF were identified for one item from the Synonyms subtest and three items from the Antonyms subtest.

Keywords: verbal ability; Woodcock-Johnson IV COG; sociocultural influence; intelligence testing; Romani children

---

TOMÁŠ MRHÁLEK, https://orcid.org/0000-0003-3872-930X; ALENA HRICOVÁ, https://orcid.org/0000-0003-0549-4962; STANISLAV ONDRÁŠEK, https://orcid.org/0000-0003-4716-3600. Correspondence concerning this article should be addressed to Alena Hricová, Zdravotně sociální fakulta, Jihočeská univerzita v Českých Budějovicích, 37005 České Budějovice, Czech Republic; e-mail: ali.kajanova@email.cz.

The paper is part of the TAČR project TL2000187 (standardisation of the Woodcock-Johnson IV for the population of Romani children). Datasets generated for this study are available from the corresponding author upon request.

Handling editors: MARTA BIAŁECKA PIKUL, Nicolaus Copernicus University, Toruń, and MAŁGORZATA PUCHALSKA-WASYL, John Paul II Catholic University of Lublin. Received 17 July 2022. Received in revised form 5 Oct. 2023. Accepted 8 April 2024. Published online 9 May 2024.

The continuous update of psychodiagnostics tools forms the basis for their effective utilization in psychological and social services. In the Czech context, new methods primarily pertain to the standardisation of internationally employed instruments, necessitating a reassessment of translation reliability. This assessment must encompass various demographic groups within the population. Therefore, this paper examines the diagnostic potential of the WJ IV COG verbal task for a group of Romani students, assessing whether its Czech adaptation is devoid of systematic measurement errors that may arise when testing individuals from non-majority populations. Specifically, our focus is on the item analysis of two fundamental subtests (Synonyms and Antonyms) using the DIF model, in accordance with the original American standardisation.

The results concentrate on the disparities between Czech norms and a group of Romani children aged 7–11 years ($N = 399$), with a particular emphasis on a comparison ($N = 204/195$) based on residence in a socially excluded locality (Čada et al., 2015) as an indicator of lower sociocultural status of the family. The Woodcock-Johnson Test of Cognitive Abilities in its fourth revision (WJ IV COG) was used as the research instrument and the Vocabulary subtest, containing the Synonyms and Antonyms subtests, was selected for the analysis presented. The paper is a partial output of the TAČR project TL2000187—Standardisation of the WJ IV (Woodcock-Johnson IV) for the population of Romani children, which focuses on the psychometric properties of this instrument in diagnosing the target group.

The objective of this study is to identify cultural loading within the verbal items of the WJ IV COG subtests and describe individual items in terms of their differential item function (DIF). Based on this, a main research question was formulated in Q1: *What items exhibit different DIF between the research sample of Romani children and the Czech standardisation dataset?*

A secondary objective was to explore the extent to which social exclusion influences the differential functioning of test items. The goal was to uncover the differences between Romani children from socially excluded localities and those living outside of these areas. The secondary research question was formulated in Q2: *Which items display different DIFs between the research sample of Romani children from socially excluded and non-excluded localities?*

Theoretical Background

A persistent issue within psychometrics, spanning various contexts, is the ongoing debate over the reliability of tests of cognitive ability for the general

population (Michell, 1997). Reynolds (1982) argued that tests are closely related to the culture in which they are developed, and hence they are culturally loaded, resulting in differential performance between individuals who have different ethnic backgrounds. The original belief was based on the notion that tests of a verbal nature are particularly subject to cultural bias, whereas tests of a non-verbal nature (e.g. those related to mathematical ability and concept formation) are not subject to cultural bias.

According to Weiss and Saklofske (2020), ethnicity can only be used as a proxy for other independent variables that affect cognitive ability test scores. Nisbett et al. (2012) emphasise that attributing variations in the performance of ethnic minority groups on intelligence tests to social and environmental factors is more appropriate, given the absence of evidence supporting the theory of genetic polymorphism. The primary criticism directed at cognitive ability tests, particularly when applied to minority populations, is based on the fact that most such tests are developed within a European-American or "Western" context, lacking consideration for the specifics of other cultures (Thaler et al., 2015).

Thaler et al. (2015) conducted a review study analysing the results of the WAIS-III cognitive ability measure. The research demonstrates that for US Hispanic and African American minorities, the differences are not just on the verbal WAIS-III, but across all subtests. In addition, the "building blocks" test is subject to cultural bias, and the authors note that low scores on arithmetic ability are more likely to reflect differences in access to education. It is therefore necessary to consider all tests as potentially subject to cultural bias, and not only tests of a verbal nature, where the difference in scores between minorities and the majority society is most apparent.

Weiss and Saklofske (2020) identified socioeconomic status (SES), of which income and education are indicators, as the main variable influencing differences in intelligence testing, and which more accurately reflects variance in IQ scores than ethnicity. It is the explanation of variance in IQ scores that is significantly reduced when it is considered, but SES predominantly affects differences in IQ scores in children, not so much in adults (Mendoza et al., 2021).

Engelhardt et al. (2019) explain differences in cognitive ability by the so-called shared environmental influences, such as parental SES, school demographics, and SES of the neighbourhood in which the child lives. These environmental influences, according to the authors of that study, affect IQ 79 percent in "verbal comprehension" and "reading skills" but only 50 percent in "maths skills". The conclusion is that parental SES is one of the most significant predictors of children's school achievement and cognitive ability. Other factors, such as race and ethnicity,

were not as influential in terms of incremental variance. Other research on intelligence testing of minorities in different countries shows the substantial influence of other factors such as socially disadvantaged backgrounds, crowded housing, environments with low cognitive stimulation (Rushton et al., 2007) or lower emphasis on regular school attendance and associated absenteeism (Dutton, 2014).

Cakirpaloglu and Kořínek (2014) define another socially disadvantaged environment as one in which there is a persistent gap between the optimal development needed for success in school and the current state of a child's abilities and capabilities, which does not offer as many developmental stimuli as are available to a child growing up in a nondisadvantaged environment. Gaertner and Tellegen (2008) administered the SON-R nonverbal test to a research sample of Romani children aged 6–12 from Amsterdam, Bratislava, and Košice, including Romani children who were adopted by parents from majority society. Children from Romani families from all countries scored below average on all four scales used with respect to majority population norms, while majority society-adopted Romani children scored average. From a qualitative perspective, the authors identified factors likely to influence cognitive ability tests in children from socially disadvantaged backgrounds—absenteeism, early school leave, language barriers, motivation, concentration, and family situation, which mainly refers to parents' value orientation related to low support in education-related activities.

Ferjenčík et al. (2015) emphasise the need for closer examination of the process of testing cognitive abilities in Romani children, especially those who come from culturally disadvantaged backgrounds. They focused their study on the measurement of cognitive functions of three groups of Romani children using the D-KEFS test battery and warned of the danger of adopting measurement tools from one sociocultural environment to another without further validation, which is particularly evident in tests containing verbal tasks. The Slovak translation of this test was standardised on 250 Slovak children aged 9–11 and 50 Romani children from socially disadvantaged backgrounds. The results of the standardisation showed systematically lower average scores for Roma children from socially disadvantaged backgrounds on all nine scales, with the most significant differences between Romani and non-Romani children in tests of a verbal nature.

Dolean et al. (2019) conducted their study in Romania on 322 Romani children aged 7–9 and growing up in a low SES environment. They concluded that socioeconomic status is a significant predictor of children's development of reading

skills and intellectual abilities, compared to a control group of 178 Romani and non-Romani children of the same age growing up in an environment characterised by high SES. Romani children from low SES environments performed significantly worse on all tests. The authors found that in addition to affecting cognitive abilities and reading skills, low SES also contributes significantly to increased absenteeism, and increased absenteeism is another factor affecting a child's cognitive abilities.

Kassis (2020) discusses the causes of low school achievement among Romani children in Greece. Segregation of Romani children and stereotyping of these children by teachers have been identified as a partial significant factor in school failure. However, it highlights the significant linguistic factor, i.e. the lack of acquisition of the language of the majority population. Bilingualism is still frequently encountered among Romani children, where in many cases neither language is dominant (the language of their family contains elements of the language of the majority population).

To address cultural biases, it is advisable to include minority groups in the normative sample. In the case of the Woodcock-Johnson Test of cognitive abilities (WJ IV COG), its standardization (McGrew et al., 2014) incorporates a relatively large normative set of 7,416 individuals chosen through quota criteria, with proportionate representation of American minority groups that are ethnically diverse. Numerous minority groups in the USA significantly impact overall standards due to their population size. For instance, the non-Hispanic black population constitutes 12.5% of the US population, while the white Hispanic population comprises 16.6% (according to the US test's technical manual). Additionally, standards consider very small minority groups with populations of less than 1%, such as "American Indian or Alaska Native" (AIANAT) or "Asian, Native Hawaiian or Other Pacific Islander" (ASIPAC) (McGrew et al., 2014). In contrast, the Czech normative sample (Furman et al., 2019) comprises respondents exclusively from the majority population due to increased collection requirements. Therefore, a supplementary analysis of the cultural independence is essential for its appropriate diagnostic application.

In contemporary psychometrics, item analysis is a central method for analysing cultural bias in a psychodiagnostics tools. Item analysis directly assesses the adjacency of a single question (i.e. item) to the trait being measured. For cognitive ability tests, in line with Rasch's model, an item's fundamental characteristic is its difficulty level, represented by the intercept that marks the ability value at which there is a 50% likelihood that the respondent will answer the question correctly at their specific ability level. On the basis of this principle,

tests can be constructed to enhance the efficiency of the diagnostic process. Each item corresponds to a level of the unidimensional construct targeted by the test, allowing empirical validation (Bond et al., 2020). This validation facilitates measuring a model of item difficulty and respondent traits along a continuum. Information about an item's relative difficulty to an individual's overall ability is useful when assessing the fit of psychological instruments within the general population and their psychometric adequacy for application across minority groups.

METHOD

Participants and Procedure

The empirical objective of this study is to conduct an item analysis on the verbal subtest of the Woodcock Test of Cognitive Ability (WJ IV COG), focused on the distortion of item functioning within this test for nonmajority populations. Czech standardisation norms were employed as a basis for comparison against the research group to examine the impact of cultural biases on the test.

The research sample consisted of Romani children ($N = 399$) between 7 and 11 years of age. The targeted subtrait under scrutiny was social exclusion, specifically related to residing in socially excluded localities. This included 204 Romani children living in such areas, according to the Analysis of Socially Excluded Localities by Čada et al. (2015), along with 195 children living outside of these localities. A comparative subset for analysis was the Czech majority population ($N = 936$), used for the Czech standardisation norms of WJ IV COG (Furman et al., 2019).

The research tool used was the Woodcock-Johnson IV COG (McGrew et al., 2014), which rests on the theoretical foundation of the CHC theory of intelligence, built on the concept of a three-layered hierarchy of abilities and relies on developmental models of cognitive ability, neurocognitive analysis, and heritability research (Abu-Hamour et al., 2012). The test consists of 18 subtests, of which, for the purposes of this paper, we work with the Verbal Vocabulary test, comprising subtests A Synonyms and B Antonyms, in the version used during the standardisation of this instrument.

Statistical analysis was performed with Winstep software (Linacre, 2012) using differential item functioning (DIF). DIF, calculated from a Rasch one-parameter model, examines the item function of the difficulty parameter for

multiple comparison groups while holding the overall level of a given ability constant between the two groups; the lower the level of DIF, the lower the item's difficulty function. As outlined by Holland and Wainer (1993), DIF analysis permits a suitable evaluation of true differences in the measured trait. This is achieved by comparing the performance on a test item between a reference group and a marginalised group, all the while controlling for overall test performance across all respondents. The results report significant differences in item difficulty between the groups based on a Rasch–Welsch *t*-test with additional Bonferroni correction to eliminate Type I error at the chosen significance level. Additional information about the level of difference based on the DIF contrast value is based on logit-level differences between the two groups, as an indicator of "effect size". Consistent with Linacre (2012), we consider a minimum level of difference to be 0.5 logits, and we separately report items greater than 1 logit.

RESULTS

Table 1 shows the results of the analysis of the Synonyms subtest, looking for differences between the children of the Romani and majority population, where differences in item difficulty of the items were found for 14 items between the Romani and the majority population. Differences were found within items 1. automobil [car] ($t = –5.01$, $p < 0.01$), 6. malý [small] ($t = –7.75$, $p < 0.01$), 11. utajit [keep secret] ($t = –2.11$, $p < 0.05$), 17. lesk [shine] ($t = –2.22$, $p < 0.05$), and 20. zdřímnout [take a nap] ($t = –3.15$, $p < 0.01$), where the DIF contrast indicated a lower item difficulty function for the Romani children. In contrast, the majority population had lower DIF in the words 2. hledět [look] ($t = 5.13$, $p < 0.01$), 3. hihňat se [giggle] ($t = 2.98$, $p < 0.01$), 4. krásný [beautiful] ($t = 0.42$, $p < 0.01$), 7. zhltnout [gulp] ($t = 2.94$, $p < 0.01$), 8. palouk [glade] ($t = 4.62$, $p < 0.01$), 10. část [part] ($t = 3.50$, $p < 0.01$), 14. zřejmý [obvious] ($t = 3.12$, $p < 0.01$), 15. kapitulovat [capitulate] ($t = 2.24$, $p < 0.05$), and 16. ruina [ruin] ($t = 2.60$, $p < 0.01$).

Table 1

Differential Item Analysis of Synonyms Versus Majority

Item	DIF Romani	DIF majority	DIF contrast	Rasch–Welch *t*	*p*-value
1. automobil [car]	**–6.45**	**–5.10**	**–1.36**	**–5.01**	**< 0.001****
2. hledět [look]	**–3.75**	**–5.01**	**1.26**	**5.13**	**< 0.001****
3. hihňat se [giggle]	**–3.97**	**–4.66**	**0.70**	**2.98**	**0.003**
4. krásný [beautiful]	**–4.92**	**–3.82**	**1.10**	**0.42**	**< 0.001****
5. rozzlobený [angry]	–3.70	–3.37	–0.33	–1.66	0.098
6. malý [small]	**–2.33**	**–0.91**	**1.42**	**–7.75**	**< 0.001****
7. zhltnout [gulp]	**–2.41**	**–3.02**	**0.61**	**2.94**	**0.003**
8. palouk [glade]	**–0.01**	**–1.23**	**1.23**	**4.62**	**0.001****
9. blyštivý [glittering]	–2.26	–2.13	–0.13	–0.62	0.537
10. část [part]	**–1.02**	**–1.86**	**0.84**	**3.50**	**0.001***
11. utajit [keep secret]	**–1.33**	**–0.81**	**–0.52**	**–2.11**	**0.036**
12. pokus [attempt]	–0.49	–0.27	–0.21	–0.77	0.440
13. tichý [silent]	–0.70	–0.15	–0.55	–1.94	0.055
14. zřejmý [obvious]	**1.83**	**0.14**	**1.68**	**3.12**	**0.003***
15. kapitulovat [capitulate]	**1.67**	**0.45**	**1.22**	**2.24**	**0.029**
16. ruina [ruin]	**2.14**	**0.20**	**1.94**	**2.60**	**0.013**
17. lesk [shine]	**–0.49**	**0.43**	**–0.92**	**–2.22**	**0.034**
18. ulice [street]	1.05	0.85	0.20	0.35	0.726
19. sužovat [harass]	3.18	1.18	2.00	1.38	0.188
20.zdřímnout si [take a nap]	**–0.83**	**1.15**	**–1.98**	**–3.15**	**0.009**

Note. Items are in bold where a statistically significant difference was found ($p < 0.05$). Bonferroni correction for repeated measures: *$p < 0.05$, **$p < 0.01$.

Table 2 shows the results of the Synonyms by Social Exclusion subtest. In terms of social exclusion, a difference was found only for the word rozzlobený

[angry] ($t = –2.49$, $p < 0.01$), in which case the item was less difficult for excluded children, while no statistically significant differences were found for the other words.

Table 2

Differential Item Analysis of the Synonyms Section by Social Exclusion

Item	DIF non-excluded	DIF excluded	DIF contrast	Rasch–Welch *t*	*p*-value
1. automobil [car]	–5.56	–6.09	0.53	1.50	0.135
2. hledět [look]	–3.00	–2.95	–0.51	–0.18	0.857
3. hihňat se [giggle]	–2.93	–3.42	0.49	1.74	0.083
4. krásný [beautiful]	–4.16	–4.28	0.12	0.42	0.674
5. rozzlobený [angry]	**–3.30**	**–2.58**	**–0.72**	**–2.49**	**0.013**
6. malý [small]	–1.45	–1.42	–0.02	–0.06	0.949
7. zhltnout [gulp]	–1.35	–1.66	0.30	0.93	0.352
8. palouk [glade]	1.12	1.33	–0.21	–0.39	0.700
9. blyštivý [glittering]	–1.52	–1.13	–0.39	–1.06	0.291
10. část [part]	0.03	0.14	–0.11	–0.25	0.806
11. utajit [keep secret]	–0.17	–0.32	0.15	0.30	0.763
12. pokus [attempt]	0.82	0.58	0.25	0.44	0.663
13. tichý [silent]	0.66	0.27	0.39	0.66	0.512
14. zřejmý [obvious]	2.91	4.28	–1.38	–0.87	0.390
15. kapitulovat [capitulate]	3.57	2.56	1.01	0.90	0.372
16. ruina [ruin]	3.99	3.34	0.65	0.43	0.673
17. lesk [shine]	0.36	1.99	–1.63	–1.60	0.125
18. ulice [street]	2.12	3.61	–1.49	–0.91	0.387
19. sužovat [harass]	4.84	4.84	0.00	0.00	1.000
20. zdřímnout si [take a nap]	0.48	1.11	–0.63	–0.40	0.717

Note. Items are in bold where a statistically significant difference was found ($p < 0.05$).

In Table 3, we find an analysis of differences in the antonyms section, which shows differences in DIF contrast (Rasch–Welch *t*-test less than 0.05) in 10 items when comparing between Romani and majority and in three cases based on residence in a socially excluded locality.

Table 3

Differential Item Analysis of the Antonym Part Compared to the Majority

Item	DIF Romani	DIF majority	DIF contrast	Rasch–Welch *t*	*p*-value
1. velký [big]	–8.90	–9.25	0.35	0.58	0.559
2. vpředu [ahead]	**–6.34**	**–7.48**	**1.14**	**2.96**	**0.003**
3. měkký [soft]	–5.87	–5.82	0.05	–0.15	0.879
4. sedět [sit]	–5.02	–4.85	–0.17	–0.60	0.551
5. bratr [brother]	**–5.40**	**–4.03**	**–1.37**	**–5.02**	**< 0.001****
6. mokrý [wet]	–4.92	–4.80	–0.12	–0.41	0.680
7. dobrý [good]	–4.24	–4.03	–0.20	–0.79	0.428
8. brečet [cry]	**–2.35**	**–1.76**	**–0.59**	**–2.98**	**0.003**
9. mlčet [be silent]	**–2.93**	**–1.97**	**–0.96**	**–4.72**	**< 0.001****
10. levný [cheap]	**–3.97**	**–3.11**	**–0.86**	**–3.60**	**< 0.001****
11. cíl [goal]	**–1.08**	**–1.80**	**0.73**	**3.55**	**< 0.001****
12. přidat [add]	–1.27	–1.25	–0.02	–0.10	0.918
13. čistit [clean]	–0.99	–0.67	–0.32	–1.63	0.104
14. líný [lazy]	–0.16	–0.24	0.08	0.39	0.699
15. otázka [question]	0.45	–0.48	0.03	0.15	0.878
16. koupit [buy]	**0.93**	**0.17**	**0.76**	**3.11**	**0.002***
17. budovat [build]	**1.31**	**0.08**	**1.22**	**4.36**	**< 0.001****
18. optimistický [optimistic]	**3.87**	**1.25**	**2.63**	**4.24**	**< 0.001****
19. pozdější [later]	**0.67**	**0.09**	**0.58**	**2.00**	**0.047**
20. bystrý [bright]	1.19	1.51	0.32	1.02	0.310

Note. Items are in bold where a statistically significant difference was found ($p < 0.05$). Bonferroni correction for repeated measures: * $p < 0.05$, ** $p < 0.01$.

The analysis of the differences in the antonyms section shows differences in item difficulty between the majority population and the sample of Romani children for items 5. bratr [brother] ($t = –5.02$, $p < 0.01$), 8. brečet [cry] ($t = –2.98$, $p < 0.01$), 9. mlčet [be silent] ($t = –4.72$, $p < 0.01$), and 10. levný [cheap] ($t = –3.60$, $p < 0.01$); it also shows a statistically significant DIF contrast in favour of the group of Romani children. On the other hand, for the population, the lower DIF contrast is found in items 2. vpředu [ahead] ($t = 2.96$, $p < 0.01$), 11. cíl [goal] ($t = 3.55$, $p < 0.01$), 16. koupit [buy] ($t = 3.11$, $p < 0.01$), 17. budovat [build] ($t = 4.36$, $p < 0.01$), 18. optimistický [optimistic] ($t = 4.24$, $p < 0.01$), and 19. pozdější [later] ($t = 2.00$, $p < 0.05$).

In Table 4 we can see a comparison between socially excluded and non-excluded Romani children, where statistically significant differences were identified for items 4. sedět [sit] ($t = –1.98$, $p < 0.05$) and 13. čistit [clean] ($t = 2.46$, $p < 0.01$) compared to item 8. brečet [cry] ($t = 2.85$, $p < 0.05$).

Table 4

Differential Item Analysis of the Antonyms Part According to Social Exclusion

Item	DIF non-excluded	DIF excluded	DIF contrast	Rasch–Welch *t*	*p*-value
1. velký [big]	–7.97	–7.97	0.00	0.00	1
2. vpředu [ahead]	–5.23	–5.70	0.47	1.07	0.286
3. měkký [soft]	–4.94	–5.07	0.13	0.33	0.742
4. sedět [sit]	**–4.60**	**–3.87**	**–0.73**	**–1.98**	**0.048**
5. bratr [brother]	–4.35	–4.70	0.34	0.91	0.365
6. mokrý [wet]	–4.23	–3.92	–0.31	–0.86	0.392
7. dobrý [good]	–3.32	–3.32	0.00	0.00	1.000
8. brečet [cry]	**–0.78**	**–1.61**	**0.84**	**2.85**	**0.005**
9. mlčet [be silent]	–2.16	–1.62	–0.54	–1.76	0.079
10. levný [cheap]	–3.10	–2.94	–0.15	–0.45	0.651
11. cíl [goal]	0.54	0.07	0.47	1.48	0.140
12. přidat [add]	0.24	–0.05	0.30	0.93	0.352
13. čistit [clean]	**0.06**	**0.91**	**–0.84**	**–2.46**	**0.014**
14. líný [lazy]	1.49	1.52	–0.03	–0.08	0.937

Table 4 (continued)

Differential Item Analysis of the Antonyms Part According to Social Exclusion

Item	DIF non-excluded	DIF excluded	DIF contrast	Rasch–Welch *t*	*p*-value
15. otázka [question]	1.16	1.16	0.00	0.00	1
16. koupit [buy]	2.85	3.08	–0.23	–0.46	0.645
17. budovat [build]	3.70	3.33	0.38	0.65	0.518
18. optimistický [optimistic]	7.12	6.40	0.73	0.49	0.625
19. pozdější [later]	2.99	2.74	0.25	0.41	0.692
20. bystrý [bright]	3.31	4.13	–0.83	–1.15	0.255

Note. Items are in bold where a statistically significant difference was found ($p < 0.05$).

DISCUSSION

The research aimed to assess which items exhibit different DIF between the research sample of Romani children and the Czech standardisation dataset. The results revealed significant discrepancies in the DIF for the group of Romani children compared to the test norms. When comparing item performance between the majority and sample groups, notable differences were observed in 14 items (70%) of the Synonyms subtest, among them 9 have a larger "effect size" than 1 logit. Data from the Antonyms subtest differed in 9 items (45%) with 6 items exceeding 1 logit. The relative difference is partially balanced, particularly within the Antonyms subtest, with 4 out of 9 Antonyms items being comparatively easier (i.e., displaying lower item difficulty) for Romani respondents, but only 4 out of the 14 Synonyms items.

Items such as "hledět" [stare] and "hihňat se" [giggle] showed a prominent increase in item difficulty for Romani children. Additionally, the words "malý" [small] and "zhltnout" [gulp] were also easier for the majority children. Significant differences were also noted for the words "palouk" [glade], "část" [part], "zřejmý" [obvious], "kapitulovat" [capitulate], and "ruina" [ruin]. These findings align with the thesis proposed by Ferjenčík, Slavkovská, and Kresila (2015) and highlight significant differences in the verbal test data. The results suggest that this specific task is influenced by the language habits and proficiency of the Romani group, which is comparatively worse compared to the majority population.

A secondary aim was to explore which items exhibit different DIF between the research sample of Romani children from socially excluded and non-excluded localities. The goal was to investigate to what extent social exclusion impacts the differential functioning of the test items. This aimed to uncover differences between Romani children originating from socially excluded localities and those living outside of them. The comparison was conducted based on the socially excluded localities identified by Čada et al. (2015).

The results obtained show several differences between living in a socially excluded area and not, specifically in one item from the Synonyms subtest of the three items from the Antonyms subtest. This may reflect the influence of the environment in which the child grows up, but given the generally low socioeconomic status of the overall research sample, these findings cannot be taken as a general effect of SES, but rather reflect the fact that singling out certain locations as socially excluded fails to capture the overall socioeconomic characteristics of the family that influence children's cognitive development. Thus, in a broader perspective, it is more effective to focus on the identification of socially disadvantaged environments as defined by Cakirpaloglu and Kořínek (2014), Rushton et al. (2007) or Nisbett et al. (2012), respectively, or by the socioeconomic status of families, rather than on social exclusion by place of residence.

A limitation of the research lies in its exclusive focus on Romani children of a younger school age, without the inclusion of older respondents from this minority group. The study focuses on this specific age group as they are most significantly affected by a possible misdiagnosis of cognitive abilities. Other intervening variables also influence cognitive performance. The analysis delves into partial socioeconomic influences represented by life in socially excluded localities, which highlight further variations in individual test items. However, living in a socially excluded locality, stated as such by the Czech Ministry of Labor and Social Affairs based on field research (Čada et al., 2015), might not necessarily represent the socioeconomic situation of a specific Romani family. The disparities found between the majority and the research sample can also be influenced by the sample size or heterogeneity.

In addition, individual family differences come into play in the research. Bilingualism within families also exerts a significant influence. Although the children primarily communicated in Czech, the linguistic environment might still impact their performance. Ferjenčík (2018) also emphasises that Romani children often lack a single fixed language and are influenced by family language habits. In the Czech Republic, Roma predominantly speak Czech, even at home. While fluent Romani usage is limited to a small number of

families, it is not uncommon to encounter an ethnolect—a blend of elements from multiple languages. For the Romani minority, this may include not only Romani but also Slovak or Hungarian.

This study examines verbal tests among Romani children aged 7–11 in comparison to Czech majority norms. Gaertner and Tellegen (2008) found language barriers as a factor contributing to test difficulties. The results show differences in item relative difficulty between groups on verbal tests, which could potentially lead to worse results in cognitive tests.

The findings also support the significance that the different linguistic background of Romani children may have, e.g. difficulties in comprehending instruction in educational settings, possibly resulting in academic underachievement. Kassis (2020) sees language barriers as one of the key determinants of school failure. Therefore, understanding the linguistic capabilities of these children is vital, so the complete list of words used as test items is included in the results with its DIFs. This information can serve for further linguistic analysis and identification of language-related challenges within the examined group.

CONCLUSION

The item analysis of the Verbal Vocabulary subtest provides insights into the differences between Romani pupils and population norms in terms of item difficulty within the Synonyms and Antonyms verbal subtests. The results reveal significant disparities in the item difficulty of the verbal cognitive abilities test for the group of Romani children aged 7–11 years, in comparison to the Czech normative (majority) population used for Czech standardisation.

Within the Synonyms section, discrepancies in item difficulty between the majority and the research group were identified in fourteen out of the twenty items examined. Among these, ten items showed higher difficulty levels for the group of Romani children, while four items showed lower difficulty. In the Antonyms subtest, differences were observed between these groups in ten items, with six items being more challenging for the research group and four for the majority.

The significance of this research lies in pinpointing these specific items to mitigate intercultural intervening influences when assessing the cognitive performance of Romani children. Moreover, the findings provide valuable insights into the linguistic habits and sociocultural knowledge of Romani children in their preschool and early school years. This discovery is crucial for evaluating

the psychometric quality of diagnostics in this group, which could impact the practical implementation of educational, health, and social interventions.

CRediT Author Statement

TOMÁŠ MRHÁLEK (33%): conceptualization, methodology, data analysis.
ALENA HRICOVÁ (33%): formal analysis, supervision, writing (original draft).
STANISLAV ONDRÁŠEK (33%): resources, writing.

REFERENCES

Abu-Hamour, B., Al Hmouz, H., Mattar, J., & Muhaidat, M. (2012). The use of Woodcock-Johnson test for identifying students with special needs – a comprehensive literature review. *Procedia – Social and Behavioral Sciences*, *47*, 665–673. https://doi.org/10.1016/j.sbspro.2012.06.714

Bond, T. G., Yan, Z., & Heene, M. (2020). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.* Routledge.

Cakirpaloglu, P., & Kořínek, R. (2014). Preschool education and school maturity of children from socially disadvantaged backgrounds. *Psychology and its Contexts*, *5*, 3–16.

Čada, K., Büchlerová, D., Kopecká, Z., Samec, T., Ouředníček, M., Sporková, J. B., Tinlová, J., Gabal, I., & Bělinová, E. (2015). *Analysis of Socially Excluded Localities in the Czech Republic*. GAC.

Dolean, D., Melby-Lervag, M., Tincas, I., Damsa, C., & Lervag, A. (2019). Achievement gap: Socioeconomic status affects reading development beyond language and cognition in children facing poverty. *Learning and Instruction*, *63*, 101–218. https://doi.org/10.1016/j.learninstruc.2019.101218

Dutton, E. (2014). Cognitive abilities among the Sami: A comment. *Intelligence*, *46*, 188–191. https://doi.org/10.1016/j.intell.2015.04.009

Engelhardt, L. E., Church, J. A., Harden, K. P., & Tucker-Drob, E. M. (2019). Accounting for the shared environment in cognitive abilities and academic achievement with measured socioecological context. *Wiley Developmental Science*, *22*(1), e12699. https://doi.org/10.1111/desc.12699

Ferjenčík, J. (2018). Diagnostika rómskych detí zo sociálne Znevýhodňujúceho prostredia [Diagnosis of Roma children from socially disadvantaged backgrounds]. *Psychológia a patopsychológia dieťaťa*, *52*(2), 127–139.

Ferjenčík, J., Slavkovská, M., & Kresila, J. (2015). Executive functioning in three groups of pupils in D-KEFS: Selected issues in adapting the test battery for Slovakia. *Journal of Pedagogy*, *6*(1), 73–92. https://doi.org/10.1515/jped-2015-0005

Furman, A., Krejčová, L., Laciga, J., Urbánek, T., & Víchová, M. (2019). *Woodcock-Johnson IV: Testy kognitivních schopností* [Woodcock-Johnson IV: Tests of cognitive abilities]. Propsyco.

Graves, S., ParkeE., Etcoff, L., San Miguel, L., & Allen, D. (2016). The relationship between the Woodcock-Johnson-III and the Batteria-III in children with attention deficit/ hyperactivity disorder (ADHD) and learning disorders. *Archives of Clinical Neuropsychology*, *31*, 584–673.

Holland, P. W., & Wainer, H. E. (1993). *Differential Item Functioning*. Lawrence Erlbaum.

Kassis, D. (2020). Underachievement of Roma Children in Greece. In E. Skourtou, V. Kourtis-Kazoullis, T. Aravossitas & P. Trifonas (Eds.), *Language diversity in Greece: Local challenges with international implications* (pp. 183–192). Language Diversity.

Linacre, J. M. (2012). *WINSTEPS* (Version 3.74.0). Chicago, IL. https://www.winsteps.com.

Mendoza, C. F., Ardila, R., Gallegos, M., & Reategui-Colareta, N. (2021). General intelligence and socioeconomic status as strong predictors of student performance in Latin American schools: Evidence from PISA items. *Frontiers in Education*, *6*. https://doi.org/10.3389/feduc.2021.632289

Michell, J. (1997). Quantitative science and definition of *measurement* in psychology. *British Journal of Psychology*, *88*, 355–383. https://doi.org/10.1111/j.2044-8295.1997.tb02641.x

Nisbett, R. E., Aronson J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*(2), 130–159. https://doi.org/10.1037/a0026699

Reynolds, C. R. (1982). The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), *The Handbook of School Psychology* (pp. 178–209). Wiley.

Rushton, J. P., Čvorović, J., & Bons, T. A. (2007). General mental ability in South Asians: Data from three Roma (Gypsy) communities in Serbia. *Intelligence*, *35*(1), 1–12. https://doi.org/10.1016/j.intell.2006.09.002

Thaler, N. S., Thames, A. D., Cagigas, X., & Norman, M. A. (2015). IQ Testing and the African American client. In L. T. Benuto & B. D. Leany (Eds.), *Guide to Psychological Assessment with African Americans* (pp. 79–80). Springer.

Weiss, L. G., & Saklofske, D. H. (2020). Mediators of IQ test score differences across racial and ethnic groups: The case for environmental and social justice. *Personality and Individual Differences*, *161*, 109–962. https://doi.org/10.1016/j.paid.2020.109962