

SLOVENSKÁ ADAPTÁCIA TESTU PRE IDENTIFIKÁCIU MATEMATICKY NADANÝCH DETÍ (TIM³⁻⁵): MOŽNOSTI A OBMEDZENIA POUŽITIA V PRAXI

MARTINA ŠINTÁLOVÁ^{1,2}, HYNEK CÍGLER¹, MICHAL JABŮREK¹, ŠÁRKA PORTEŠOVÁ¹, ONDŘEJ STRAKA¹

^{1,2} *Institut pro psychologický výzkum, Fakulta sociálních studií, MU, Brno*

² *Centrum dopravního výzkumu, v. v. i., Brno*

ABSTRACT

Objectives. The number of identified gifted children in Slovakia is small and Slovak psychologists need more modern and specialized assessment tools. Test for Identifying Gifted Children in Mathematics in Grades 3–5 (TIM³⁻⁵) is standardized in Czechia and has excellent psychometric properties with two parallel and fully equated forms. The study aims to adapt the test in the Slovak language and support its practical use. The authors ran a small pilot study using form A only and compared it to the Czech standardization sample. **Sample and settings.** The Slovak sample consisted of 169 pupils from four elementary schools and was supplemented by the Czech standardization sample (404 children in form A). The Slovak data were collected in February 2022; informed consent from the parents/legal guardians was obtained before data collection with approx. 40 % drop-out. The original authors of the test provided the Czech data collected in 2015. **Hypotheses and analysis.** The authors compared descriptives of Slovak and Czech pupils, assessed essential psychometric parameters (reliability and factor validity), and mainly performed measurement invariance and Differential Item Functioning analyses. **Results.** The psychometric parameters of the Slovak test form were excellent, fully comparable, or even better than the original Czech version. The mean of IRT reliability across grades was .76. However, the test differentiates mainly

in above-average children. The confirmatory IRT analyses suggested clear unidimensionality and scalar invariance across the Czech and Slovak samples. However, Slovaks outperformed Czech pupils. The difference was highest in the third grade with 9.6 and $_{95\%}CI = [6.7-12.1]$ points at the T-score scale and lowest in the fifth grade, 3.6 with $_{95\%}CI = [0.9-6.3]$ points. The authors recommend the test for practical use using Czech norms. However, a user should be aware of possible differences in average performance, considering that the Czech norms could be too mild for Slovak children. **Limits.** Such a massive difference between Czech and Slovak populations is not plausible. The authors provided several explanations based mainly on the sampling procedure and systematic missingness in the Slovak sample, correlated to math ability. The most realistic explanation of the difference is a systematic sampling error in one or both samples. Therefore, the results are of importance for Czech users as well since it might be the case that the norms are too mild also for the Czech pupils. The authors advise Czech test users to interpret the test results with caution (and rather conservatively) until a new norming study is performed.

key words:

giftedness,
mathematical performance,
assessment of gifted children,
test adaptation,
elementary school

ÚVOD

Identifikácia mimoriadneho nadania je prvým krokom v procese vzdelávania a rozvoja nadaných detí. Vzhľadom ku komplexnosti konceptu nadania a jeho možných prejavov je nevyhnutná dôkladná diagnostika za použitia kvalitných, štandardizovaných metód. V zahraničí sú k identifikácii nadaných žiakov najčastejšie používané

Došlo: 14. 4. 2023; H. C., Institut pro psychologický výzkum, Fakulta sociálních studií, Masarykova univerzita, Joštova 10, 602 00 Brno; e-mail: cigler@fss.muni.cz

Táto štúdia bola vypracovaná v rámci projektu MUNI/A/1473/2022 a získala podporu z Grantových projektov Programu špecifického výskumu Fakulty sociálních studií Masarykovy univerzity.

inteligentné testy a tzv. testy školských schopností (*tests of achievement*; Perleth et al., 2000), ktoré sa zameriavajú najmä na jazykové a matematické schopnosti. Výhodou takýchto testov schopností je, že ich administrácia nie je viazaná na psychológa, vďaka čomu ich môžu používať aj samotní pedagógovia a nadväzovať na ne prípadnými opatreniami pri ďalšej výuke. Keďže matematika patrí spolu s jazykmi medzi najčastejšie testované oblasti, existencia kvalitných metód by nemala byť opomínaná. V otázke diagnostiky matematického nadania aktuálne na Slovensku existuje pomerne prázdne miesto, preto by adaptácia českého Testu pre identifikáciu matematicky nadaných žiakov (TIM³⁻⁵) do slovenského prostredia, ktorá je cieľom tohto výskumu, mohla pomôcť tento priestor aspoň čiastočne zaplniť.

V nasledujúcom texte zhrnieme diagnostické metódy pre diagnostiku (matematického) nadania, ktoré sú aktuálne dostupné na Slovensku a ukážeme, že ich je kritický nedostatok. Následne predstavíme test TIM³⁻⁵ a navrhne a dôkladne odôvodníme design adaptačnej štúdie. V tomto ohľade sa namiesto bežnej adaptácie (teda lokálnej štandardizačnej štúdie a tvorby slovenských noriem) zameriame na viac atypický postup, ktorý overí invarianciu merania českého originálu a slovenskej adaptácie.

Metódy posudzovania matematických schopností na Slovensku

Podľa Metodických pokynov na zaraďovanie detí do špeciálnych výchovno-vzdelávacích programov pre intelektovo nadaných žiakov (2005) by sa diagnostike nadania na Slovensku mali venovať Centrá pedagogicko-psychologického poradenstva a prevencie (CPPPaP), preto v rámci zisťovania aktuálneho stavu praxe v oblasti diagnostiky matematického nadania bolo mailom oslovených 97 CPPPaP z rôznych častí Slovenska¹. Z tohto množstva na mail odpovedalo 20 centier, pričom 18 z nich uviedlo, že sa špecificky diagnostike matematického nadania nevenujú. Z hľadiska nadania sa zaoberajú väčšinou iba diagnostikou všeobecného intelektového nadania, pričom za týmto účelom je v oslovených centrách jednoznačne najčastejšie používaná Wechslerova intelligenčná škála pre deti – WISC-III (Wechsler et al., 2006). Menej často, no v troch prípadoch bol tiež spomenutý test Woodcock Johnson International Edition – WJ IE (Ruef et al., 2003). Pomerne prekvapivé množstvo psychológov (celkovo 6) v tomto kontexte spomenulo Ravenove progresívne matice (Ferjenčík & Hromý, 1989) a Urbanov figurálny test tvorivého myslenia (Urban et al., 2002), ktorý bol spomenutý 4-krát, a test SOCAG (Dočkal, 2000), zameraný na diagnostiku sociálneho správania. V jednom alebo dvoch prípadoch boli zmienené tiež metódy ako Kohsove kocky (Košč, 2002), I-S-T 2000 (Amthauer et al., 2005), Analýza štruktúry inteligencie – ISA (Fay et al., 2001), Kaufmanova hodnotiacia batéria pre deti – K-ABC (Kaufman & Kaufman, 1999), Stanford-Binetova intelligenčná škála (Terman & Merrill, 1972), Názorové rady (Vonkomer & Miglierini, 1992) a Testy kognitívnych schopností (Thorndike & Hagen, 1998). Ako už bolo zmienené, vymenované testy sú používané za účelom diagnostiky všeobecného nadania, nie matematických schopností.

Z testov zameraných špecificky na matematické schopnosti sa objavili najčastejšie Testy matematických schopností (Košč, 1987), v rámci ktorých sú zahrnuté testy Kalkúlia III, Číselný trojúholník a Rey-Osterriethova komplexná figúra (všetky tri testy však boli viackrát zmienené aj ako samostatné metódy, nie len v kontexte tejto batérie), ďalej Farebná kalkúlia (Novák, 2001) a ZAREKI (von Aster & Weinholdová, 2008).

¹ Tento prieskum je len orientačný a slúžil iba pre účely získania všeobecného prehľadu o aktuálnom stave metód používaných na Slovensku k diagnostike matematického nadania. Získané dáta nemusia byť reprezentatívne.

Väčšina z uvedených testov je zameraná na nižšie, prípadne stredné pásmo matematických schopností, pričom diskriminácia v oblastiach vyššieho nadpriemeru môže byť sporná. Z uvedených testov by, podľa informácií v manuáli, k vyhl'adávaniu nadaných žiakov mohli slúžiť testy Farebná kalkúlia a Kalkúlia III, tie však postihujú iba malú časť matematických schopností a vzhľadom k nedostatočne preskúmaným psychometrickým vlastnostiam poskytujú len orientačný náhľad na tieto schopnosti. Vyššie spomínaná Rey-Osterriethova komplexná figúra sa taktiež ukázala ako nevhodná k diagnostike matematicky nadaných jedincov (Krčová et al., 2017).

Dalšou uvádzanou možnosťou merania matematických schopností je použitie vybraných subtestov z intelligenčných škál, avšak jednotlivé subtesty majú oproti celej testovej batérii nižšiu reliabilitu a vyššiu chybu merania, preto aj keď administrácia komplexnej intelligenčnej batérie je už bežnou súčasťou diagnostiky nadania, spoliehať sa v prípade diagnostiky matematického nadania iba na jednotlivé subtesty nie je dostačujúce (Cígler et al., 2017b). Okrem uvedených štandardizovaných metód je k posudzovaniu matematických schopností tiež možné použiť rôzne neštandardizované súbory matematických úloh a testov, ktorých reliabilita a validita však nie sú overené. Okrem toho je tiež možné, že takéto úlohy merajú najmä úroveň zvládnutia školského učiva a úroveň nadania postihujú len v menšej miere.

Z detailnej rešerše existujúcich metód, ako aj z informácií poskytnutých psychológmi z viacerých CPPPpP, je preto možné konštatovať, že aktuálne na Slovensku nie je k dispozícii žiadna štandardizovaná metóda určená primárne k identifikácii matematického nadania.

Test pre identifikáciu nadaných žiakov v matematike – TIM³⁻⁵

TIM³⁻⁵ (Cígler et al., 2017a, 2017b) bol vyvinutý tímom z Institutu pro psychologický výzkum na Fakultě sociálních studií Masarykovy univerzity so zámerom vyplniť medzeru, ktorá v tom čase v oblasti identifikácie matematicky nadaných detí v Českej republike existovala. Jedná sa o doménovo špecifický test, určený k diagnostike mierne až vysoko nadpriemerných matematických schopností, vďaka čomu je schopný eliminovať tzv. efekt stropu, teda prípad, kedy by všetci žiaci od istej úrovne schopností dosiahli maximálneho skóre a nebolo by tak možné medzi úrovňou ich schopností bližšie diferencovať (Cígler et al., 2017b). Z hľadiska CHC teórie inteligencie (Schneider & McGrew, 2012) meria úzku schopnosť matematické usudzovanie (*quantitative reasoning*), ktorá je súčasťou širšieho faktoru fluidná inteligencia. Čiastočný vplyv na výkon v teste však majú aj ďalšie charakteristiky dieťaťa, ako napríklad motivácia, vedomosti a iné, preto je možné ho považovať aj za test školského výkonu. Testy s podobnou konštrukciou bývajú aj súčasťou batérií školských zručností, vrátane testov WJ-IV Tests of Achievement (Schrank et al., 2014).

Test je tvorený 25 položkami, ktoré majú formu matematických úloh a boli zámerne vytvárané tak, aby vyžadovali čo najmenšie množstvo konkrétnych znalostí (napr. terminológie, postupov, matematickej symboliky, atď.) a minimalizovali tak vplyv tohto faktoru na výsledky. Väčšina položiek je vo forme otvorených odpovedí s dostatočným priestorom pre zápis postupu riešenia, čo umožňuje sledovanie myšlienkových postupov dieťaťa pri riešení úlohy, prípadne zohľadnenie čiastočne správneho postupu aj v prípade nesprávnej odpovedi. V rámci administrácie dieťa dostane pracovný zošit s úlohami, ktoré samostatne rieši po dobu maximálne 45 minút. Test je možné administrovať individuálne aj skupinovo a obsahuje dve paralelné formy, A a B, ktoré umožňujú opakované testovanie toho istého žiaka, prípadne znižujú pravdepodobnosť opisovania pri skupinovej administrácii. Je určený na použitie predovšetkým psychológmi, prípadne špeciálnymi pedagógmi alebo pedagógmi základných škôl a vďaka on-line vyhodnocovacej aplikácii je jeho vyhodnotenie jednoduché a rýchle.

Test obsahuje tri typy úloh – aplikačné, geometrické a aritmetické. Faktorová analýza v rámci štandardizačnej štúdie však potvrdila jednodimenzionálnu štruktúru testu. Reliabilita v zmysle vnútornej konzistencie štandardizačnej verzie sa v závislosti na ročníku a forme pohybuje od 0,66 do 0,91. V súlade s jeho zameraním test rozlišuje najmä v pásme stredného a vyššieho nadpriemeru, čo podhodnocuje odhad reliability pri žiakoch tretích ročníkov (Cígler et al., 2017a).

Validita bola dokázaná pomocou diskriminačnej účinnosti medzi nadanými a bežnými žiakmi, posúdením vzťahu výsledkov v teste so školskými známkami, hodnotením učiteľov, inteligenčnými testami WJ IE II COG (Ruef et al., 2010) a WISC-III (Wechsler et al., 2002) a ďalšími premennými. Bližšie informácie o jeho reliabilite, validite, aj spôsobe vývoja testu je možné nájsť vo voľne dostupnej publikácii (Cígler et al., 2017a).

Kultúrne a systémové rozdiely medzi ČR a SR

Pri adaptácii testu z krajiny jeho vzniku do nového prostredia je dôležité zohľadniť všetky potenciálne rozdiely medzi týmito prostrediami, ktoré by mohli ovplyvniť správne fungovanie tohto testu. Pravdepodobne najvýraznejší rozdiel medzi českým a slovenským vzdelávacím systémom na základných školách je z hľadiska organizácie výuky. Vzdelávanie na ZŠ je v ČR aj na Slovensku rozdelené na 1. a 2. stupeň. V ČR pod 1. stupeň spadajú ročníky 1.–5., pod 2. stupeň potom 6.–9. ročník. Na Slovensku sú v rámci 1. stupňa radené len ročníky 1.–4. a 5. ročník je zahrnutý už pod 2. stupeň ZŠ. Na základe zbežného porovnania minimálnych výstupných kompetencií v predmete matematika pre 3. a 5. ročník v rámci Rámcového vzdelávacieho programu pre základné vzdelávanie (RVP ZV) v ČR a Štátneho vzdelávacieho programu (ŠVP) na Slovensku by však nemal medzi týmito dvoma krajinami existovať zásadný rozdiel.

Ďalším drobným kultúrno-spoločenským rozdielom medzi Českou republikou a Slovenskom je odlišná platobná mena. Tento rozdiel by mohol byť relevantný z dôvodu, že vo viacerých úlohách testu sa nachádzajú peňažné sumy, ktorých menu bolo v rámci adaptácie nutné zmeniť. Tento aspekt je aj s možnými dôsledkami bližšie zmienený v sekcii Metóda.

V rámci medzinárodného testovania PISA (Programme for International Student Assessment) z roku 2018 (Schleicher, 2019) dosiahli v matematickej časti české deti vo veku 15–16 rokov priemerného výsledku 499 ($SD = 93$), zatiaľ čo slovenské deti 486 ($SD = 100$), čo je pri uvedených veľkostiach smerodajnej odchýlky zanedbateľný rozdiel (Cohenovo $d = 0,13$). Priemerný výsledok v testovaní matematických schopností TIMSS (Trends in International Mathematics and Science Study) z roku 2019 bol pre štvrtý ročník ZŠ v Českej republike 533, na Slovensku 510 (Mullis et al., 2020), čo so smerodajnou odchýlkou 100 opäť predstavuje len veľmi malý rozdiel. Na základe vyššie uvedených informácií je preto možné sa domnievať, že neexistuje závažný dôvod k tomu, aby sa české a slovenské deti výrazným spôsobom líšili v miere matematických schopností.

Výskumný cieľ

Vzhľadom ku zjavnému nedostatku vhodných metód pre identifikáciu nadaných detí na Slovensku je našim cieľom adaptovať test TIM³⁻⁵ do slovenského prostredia a uviesť ho do tunajšej distribúcie. Pretože „plná“ adaptácia vrátane konštrukcie noriem na rozsiahly súbore by mohla byť veľmi nákladná a teda neefektívna, zvolíme alternatívny spôsob v podobe pilotnej štúdie na menšom súbore. Na tomto súbore budú overené iba základné psychometrické parametre (prevažne reliabilita a faktoro-

vá štruktúra), porovnáme deškríptívne štatistiky slovenských a českých detí a najmä vykonáme analýzu invariance, ktorá overuje, či test meria v oboch skupinách rovnaký latentný rys. Súčasťou štúdie je tiež kvalitatívne posúdenie reakcií slovenského súboru. V prípade, že budú rozdiely slovenského pilotného a českého štandardizačného súboru zanedbateľné, bude možné použiť české normy aj pre slovenské deti.

METÓDA

Postup prekladu

Vzhľadom na značnú podobnosť českého a slovenského jazyka bol zvolený priamy preklad do slovenčiny bez nutnosti spätného prekladu (*forward translation*). Obidve formy testu boli preložené jedným z autorov štúdie, rodeným slovenským hovoriacim, pričom zrozumiteľnosť a frekvencovanosť použitia niektorých výrazov bola počas procesu prekladu konzultovaná s tromi ďalšími rodenými Slováckmi, vrátane jednej osoby vo veku cieľovej kategórie testu (10–11 rokov).

Preložené verzie testu boli následne porovnávané s pôvodnou českou verzou tromi nezávislými hodnotiteľmi z hľadiska zachovania významu, jasnosti inštrukcií, zachovania gramatických štruktúr a zachovania komplexity vyjadrenia. Kritérium zachovania významu sa vzťahovalo k zachovaniu obsahovej validity položky, v rámci jasnosti inštrukcií sa hodnotila najmä zrozumiteľnosť položky pre deti v cieľovej kategórii testu, kritérium zachovania gramatických štruktúr berie do úvahy, či zostali nezmenené napr. jednotné/množné číslo, osoba, čas a ďalšie gramatické aspekty textu. Posledné kritérium, komplexita vyjadrenia, sa týka formálnej štylizácie položky, teda napríklad vhodnosti použitých slov, dĺžka a zložitosť viet a podobne. V každom z hodnotených kritérií mohla byť každej položke pridelená hodnota 0, 1 alebo 2, kde 0 značila zachovanie, 1 čiastočné zachovanie a 2 narušenie daného kritéria. Jediné v prípade hodnotenia jasnosti inštrukcií znamenala 0 zachovanie jasnosti (bez zmeny), hodnotenie 1 označovalo jasnejšie inštrukcie v preloženej verzii a hodnotenie 2 naopak rozporupnejšie inštrukcie. V online supplimente <https://doi.org/10.17605/OSF.IO/6TFV4> je priložená tabuľka zobrazujúca všetky položky, ktoré získali v akomkoľvek kritériu hodnotenie iné ako 0. Celkovo 31 položiek (z celkových 50) bolo v niektorom z kritérií hodnotených inou hodnotou ako 0. Na základe hodnotení a slovných pripomienok bolo znenie uvedených položiek prepracované, vo väčšine prípadov sa však jednalo iba o drobné zmeny vo formuláciách položiek. Výraznou zmenou prešla jedine položka B10, ktorej pôvodný preklad nespĺňal kritérium zachovania komplexity vyjadrenia. Prvý preklad položky obsahoval oproti českej verzii navyše zátvorku a znenie bolo celkovo dlhšie. Z tohto dôvodu bola preformulovaná do jej súčasnej podoby, ktorá je bližšie zmienená nižšie v texte.

Test bol prekladaný s cieľom zachovať výsledné znenie čo najprirodzenejšie, preto v prípade, že doslovný preklad nezodpovedal štruktúre slovenského jazyka boli preferované skôr funkčné ako doslovné ekvivalenty. Napríklad, v úlohe A7 bol výraz *početní výkon* zmenený na *matematická operácia*. Aj napriek tomu, že sa nejedná o doslovný preklad, po konzultácii s rodenými Slováckmi bol výraz *matematická operácia* zhodnotený ako frekvencovanejší a zrozumiteľnejší než *počtový výkon*. Ďalšia úloha, v ktorej došlo k drobnému významovému posunu je B25. Dôvodom je, že v slovenčine neexistuje ekvivalent českého slovesa *utkat se*. Významovo najbližšie by bola pravdepodobne formulácia *stretnúť sa v boji*, čo však textu uberá na prirodzenosti. Zvolený preklad *rytier vyzval na súboj trojhlavého draka* namiesto českého *rytíř se utkal s tříhlavým drakem* preto síce mierne mení význam vety, zároveň by to však nemalo mať žiaden odôvodnený vplyv na náročnosť úlohy.

Špecifikom boli úlohy obsahujúce peňažné sumy v českej mene (A5, A9, A19, B6, B7, B10). Aby nebol narušený aspekt toho, že sa jedná o domácu menu, v slovenskej verzii boli použité eurá, ktoré sú slovenským deťom pravdepodobne najbližšie. Číselné údaje boli v úlohách zachované, keďže však hodnota v eurách nezodpovedá hodnote v českých korunách, bolo nutné pozmeniť niektoré formulácie úloh. Konkrétne napríklad v úlohe A9 bolo pôvodné slovo *skleničky* zmenené na *knihy*, nepredpokladá sa však, že by táto zmena mala mať vplyv na náročnosť úlohy. V úlohe B10 muselo pri zachovaní číselných hodnôt dôjsť k výraznejšej obsahovej zmene zadania a v procese prekladu a následného hodnotenia bola viackrát revidovaná. Pôvodná formulácia *kytice ze dvou gerber a tři karafiátů stojí 105 Kč* bola v prvej verzii zmenená na *ceny do školskej súťaže (3 lístky na hokejový zápas a 2 lístky do divadla) boli zakúpené za 105 eur*. Zahŕnutie zátvorky však znenie položky komplikovalo a potenciálne mohlo predlžovať čas nutný na porozumenie zadaniu, nakoniec preto bola položka zmenená na *farmár si kúpil na farmu zvieratká za 105 eur, celkovo tri sliepky a dve prasiatka*. Slovenská verzia úlohy je aj v aktuálnom znení o niekoľko slov dlhšia a obsahuje súvetie, čo by mohlo potenciálne mať menší vplyv na čas potrebný na prečítanie úlohy. Toto riziko berieme do úvahy, zároveň to však považujeme za lepšie riešenie ako ponechanie pôvodnej kytice s nezodpovedajúcou hodnotou. V úlohách A5, A19, B6 a B7 nedošlo v preklade okrem zmeny českých korún za eurá k ďalším zmenám, avšak je potreba brať do úvahy, že sa v slovenskej verzii pracuje s približne 25-krát vyššou nominálnou hodnotou ako v českej verzii, čo by, najmä pre mladšie deti, pre ktoré sumy v desiatkach eur nemusia byť dostupné, mohlo mierne zvýšiť náročnosť úlohy.

K zmene došlo taktiež v prípade niektorých mien, tak aby sa jednalo o mená bežnejšie v slovenskom jazyku (napr. *Jirka* na *Jarko*, *Honza* na *Jožko*).

Výskumný súbor

Slovenský výskumný súbor tvorili deti z 3. až 5. ročníka zo 4 základných škôl na Slovensku. Oslovených bolo približne 60 škôl, z ktorých sa spoluprácu podarilo nadviazať so štyrmi – v Trnave, Trenčíne a dvoma školami v Žiline. Zriaďovateľom prvých dvoch škôl je štát, zatiaľ čo v prípade tretej a štvrtej školy sa jedná o katolícke cirkevné školy, z hľadiska výuky matematiky by sa však nemali líšiť. Celkovo sa výskumu zúčastnilo 169 detí, z toho 90 dievčat (53 %) a 79 chlapcov (47 %). V tab. 1 sú uvedené bližšie informácie ohľadne zloženia výskumného súboru podľa školy, pohlavia a ročníku.

Tab. 1 Popis výskumného súboru

		ročník						n
		3		4		5		
		chlapci	dievčatá	chlapci	dievčatá	chlapci	dievčatá	
Žilina 1	cirkevná	6	4	5	4	5	5	29
Trnava	štátna	9	4	3	13	12	9	50
Trenčín	štátna	12	7	6	5	6	14	50
Žilina 2	cirkevná	3	9	6	9	6	7	40
spolu		30	24	20	31	29	35	169
		54		51		64		

Z hľadiska veku sa priemer detí medzi jednotlivými ročníkmi líšil približne o 11 mesiacov, čo zodpovedá očakávaniam. Slovenské deti boli v priemere o 3,2 mesiaca staršie ako české deti ($M = 9,9$; $SD = 0,91$), $t(334,1) = 3,25$, $p = 0,001$. Z hľadiska pohlavia boli dievčatá v rámci ročníku priemerne o 1 až 2 mesiace mladšie. Priemerný vek (uvedený vo formáte roky; mesiace) a smerodajné odchýlky (v mesiacoch) pre každý ročník sú uvedené v tab. 2.

Tab. 2 Rozdelenie respondentov podľa veku

ročník	chlapci		dievčatá		spolu	
	M	SD	M	SD	M	SD
3	9;3	5,5	9;2	4,4	9;2	5,1
4	10;2	4,9	10;0	5,0	10;1	5,0
5	11;1	4,8	10;12	3,8	11;0	4,3

Poznámka. M – priemery vo formáte rok; mesiac; SD – smerodajné odchýlky v mesiacoch

Žiadne z detí nebolo pred testovaním oficiálne diagnostikované ako nadané, čo môže mať viacero vysvetlení. Ak by sme počítali s normálnym rozložením schopností v populácii a brali do úvahy ako štandard pre nadpriemerný výkon umiestnenie dve smerodajné odchýlky od priemeru, jednalo by sa o 2,2 % detí, čo by v rámci nášho súboru mali byť 3 deti. Je však možné, že výskyt nadaných detí v testovaných triedach je nižší z dôvodu, že deti oficiálne diagnostikované ako nadané sú následne preradené do špecializovaných tried alebo škôl. Teoreticky by to ale mohlo tiež naznačovať, že diagnostika nadania na Slovensku pravdepodobne nie je súčasťou bežnej psychologicko-pedagogickej praxe. Pravdepodobne sa môže jednať o kombináciu týchto faktorov.

Známky z matematiky a slovenského jazyka boli získané u 150 detí (89 %), v tab. 3 sú uvedené početnosti a percentuálne zastúpenie jednotlivých známok podľa ročníku a pohlavia. Najčastejšou známkou celkovo a zvlášť pre 3. a 4. ročník bola 1, avšak v 5. ročníku mierne prevažovali 2 v prípade oboch predmetov.

Zbierané boli tiež údaje o špecifických poruchách učenia (ŠPU), vyskytli sa však iba dyslexia ($n = 3$) a dysgrafia ($n = 2$). Ďalšie ŠPU (ako napr. dyskalkúlia, dysortografia) neboli uvedené u žiadneho zo žiakov. Z ďalších porúch sa (vždy v jednom prípade) vyskytla porucha CNS, znížený intelekt, porucha pozornosti, ADHD a ľahká porucha autistického spektra. V prípade 3 žiakov bol uvedený individuálny výchov-

Tab. 3 Zastúpenie jednotlivých známok podľa ročníku

ročník	matematika			spolu		slovenský jazyk			spolu	
	3	4	5	n	%	3	4	5	n	%
známka	n	n	n	n	%	n	n	n	n	%
1	32	41	29	102	60	28	35	27	90	53
2	3	9	30	42	25	6	13	30	49	29
3	0	0	4	4	2,5	1	2	5	8	5
4	0	1	1	2	1,5	0	1	2	3	2
5	0	0	0	0	0	0	0	0	0	0
NA	19	0	0	19	11	19	0	0	19	11

no-vzdelávací plán (IVVP), ktorý svedčí o pravdepodobnom znevýhodnení týchto žiakov, informácie o konkrétnej diagnóze však pedagóg ani psychológ neboli schopní poskytnúť. Celkovo u 158 žiakov (93 %) nebola poskytnutá žiadna diagnóza, avšak v niekoľkých prípadoch pedagógovia uviedli, že majú podozrenie na existujúce ŠPU u konkrétnych žiakov aj napriek tomu, že im informácia o oficiálnej diagnóze nebola zo strany rodičov poskytnutá. Reálny výskyt ŠPU v skúmanom súbore preto môže byť teoreticky vyšší, ale keďže pri niektorých žiakoch nemáme informácie o oficiálnej diagnóze z Centra pedagogicko-psychologického poradenstva a prevencie (CPPPaP), berieme do úvahy iba vyššie uvedené počty. Vzhľadom k takto malému počtu detí s jednotlivými diagnózami ďalej tieto premenné nezvažujeme a pracujeme s celým súborom dohromady.

Pedagógovia na základe ich subjektívneho hodnotenia označili spolu až 41 žiakov (24 %) ako šikovných, rozdelenie podľa ročníku a pohlavia je uvedené v tab. 4. Percentuálne zastúpenie žiakov označených ako šikovných sa v závislosti od školy pohybovalo od 18 % do 30 %, pričom nie je možné posúdiť, do akej miery je tento rozdiel daný optimistickjším pohľadom pedagógov alebo skutočným rozdielom vo výskyte šikovných žiakov na jednotlivých školách.

Tab. 4 Rozdelenie „šikovných“ žiakov podľa ročníka a pohlavia

ročník	chlapci	dievčatá	spolu
3	7	4	11
4	7	12	19
5	3	8	11
spolu	17	24	41

Slovenský súbor bol ďalej doplnený o súbor českých detí, ktorý vznikol vo februári 2015 a bol použitý pri českej štandardizácii. Jedná sa o 404 detí; jeho charakteristiky a spôsob zberu dát je podrobne popísaný v online dostupnom psychometrickom manuáli (Cígler et al., 2017a).

Zber dát

Výskum bol schválený Etickou komisiou pre výskum Masarykovej univerzity (číslo jednania EKV-2021-094) v decembri 2021. Zber slovenských dát prebiehal v období od 8. 2. do 24. 2. 2022, pričom približne týždeň pred samotným zberom boli školám osobne distribuované informované súhlasy pre zákonných zástupcov detí, ktorých podpis bol podmienkou pre zapojenie dieťaťa do výskumu. Návratnosť informovaných súhlasov bola pomerne nízka (zhruba od 30 do 80 % v jednotlivých triedach, priemerne okolo 50 %), pričom nie je vylúčené, že sa jednalo o systematický úbytok.

Všetkým zapojeným žiakom bol pridelený kód, vďaka čomu bola zaistená anonymita dát z pohľadu autorov tejto štúdie. Triedni učitelia (ktorí mali ako jediní prístup k údajom žiakov) boli požiadaní o vyplnenie Záznamového listu, v rámci ktorého sa o žiakoch anonymne zbierali údaje o ich mesiaci a roku narodenia, známkach z matematiky a slovenského jazyka pri poslednom polročnom hodnotení a prípadnej diagnóze nadania, ŠPU alebo iných porúch. Učitelia mali taktiež možnosť do Záznamového listu zaznačiť žiakov, ktorých subjektívne považovali za šikovných.

K zberu dát bola využitá výhradne forma A, forma B z dôvodu dvojnásobných nárokov na veľkosť výskumného súboru administrovaná nebola. Samotná administrácia testu prebiehala vo všetkých prípadoch počas vyučovacej hodiny, pričom podmienky v jednot-

livých triedach sa mierne líšili z hľadiska počtu žiakov v triede, prítomnosti pedagóga či priestorového rozloženia. Postup administrácie bol vo všetkých prípadoch rovnaký – na začiatku hodiny sa žiakom vysvetlil účel a priebeh testovania, boli informovaní o možnosti kedykoľvek z výskumu odstúpiť a o tom, že ich výsledok v teste nebude známovaný. Časť dát (celkovo 5 tried, 59 detí) bola zozbieraná školskou psychologičkou pracujúcou na daných školách z dôvodu nepredvídateľného zatvárania niektorých tried do karantény, spojeného s pandemiou COVID-19. Pred administráciou jej bol poskytnutý presný súpis pokynov s cieľom dosiahnuť čo najpodobnejšie podmienky administrácie a po následnom spoločnom zhodnotení priebehu zberu dát nie je dôvod sa domnievať, že by došlo k výraznému odklonu od štandardnej formy administrácie.

Počas administrácie testu sa opakovane objavovali špecifické, potenciálne problematické situácie, ktoré je možné zhrnúť do nasledujúcich kategórií. Prvou sú otázky detí počas písania testu, pričom pravdepodobne najčastejšie pokladaná otázka (na základe subjektívneho hodnotenia administrátora, kvantitatívne dáta o počte otázok neboli zbierané) sa týkala možnosti opakovania číslíc v položke A8. Z hľadiska možného ovplyvnenia ostatných detí pri odpovedi na túto položku je pri skupinovej administrácii iba položenie tejto otázky pomerne problematické, možné dôsledky sú bližšie rozoberané v kapitole Diskusia. Ďalšou opakovane pokladanou otázkou bolo neporozumenie časti testu, označenej ako A12–A14, ktorá je úvodom k položkám A12 až A14. Deti sa domnievali, že sa jedná o samostatné položky a neboli si isté, čo je v tejto časti ich úlohou. Z hľadiska prípadného ovplyvnenia výsledkov túto nejasnosť nepovažujeme za zásadnú, preto sa jej ďalej nebudeme venovať. Ďalšou špecifickou kategóriou je častý nedostatok priestoru v triedach na to, aby deti sedeli samostatne. Keďže administrovaná bola iba forma A, potenciálne sa tak zvýšilo riziko prípadného odpisovania detí. Posledným potenciálnym problémom, ktorý sa počas administrácie objavoval, bol nedostatok času k dokončeniu testu. Ak vezmeme do úvahy, že rozdanie testov, poskytnutie inštrukcií a odpovedanie na prípadné otázky v jednotlivých triedach zabralo 5–10 minút a štandardná dĺžka vyučovacej hodiny na Slovensku je 45 minút, znamená to, že vo väčšine prípadov deti nemali k dispozícii celých 45 minút na písanie testu. Vo väčšine prípadov deti odovzdali testy ešte pred zvončením, ak však niektoré dieťa bolo ochotné písať aj cez prestávku a bolo to v limite 45 minút od začiatku testovania, bolo mu to umožnené.

Analýza dát

Všetky IRT modely boli odhadnuté v R balíčku *mirt* (Chalmers, 2012) pomocou „expected-maximization“ (EM) algoritmu; vo všetkých prípadoch boli odhadnuté výhradne jednodimenzionálne modely. Položky boli v zhode s originálnou verziou testu parametrizovanou pomocou raschovského partial credit modelu (PCM), resp. dvojparametrového generalized partial credit modelu (GPCM). Zhoda modelu s dátami bola overená pomocou bežných indexov dobrej zhody (*SRMSR*, *RMSEA* vrátane 90% intervalu spoľahlivosti a *TLI*) vypočítaných na základe M_2^* štatistiky (Cai & Monro, 2014). V prípade *gpcmIRT* parametrizácie položiek (viď nižšie) balíček *mirt* bohužiaľ neposkytuje odhad *SRMSR*.

Pre analýzu invariance sme využili tradičnú parametrizáciu GPCM položiek (v prostredí balíčku *mirt* označenú ako *gpcmIRT*), ktorá oddeľuje parameter obtiažnosti položky a prahy jednotlivých kategórií; analýza prebiehala v niekoľkých krokoch. V konfiguračnom modeli boli všetky parametre položiek odhadnuté pre Čechov a Slovákov zvlášť, priemer a rozptyl latentnej premennej bol fixovaný v oboch skupinách (na 0, resp. 1). V druhom kroku bola overená prahová invariancia. U ordinárnych položiek sme obmedzili prahy položiek na rovnakú hodnotu, zatiaľ čo celková

obtiaznosť položky v slovenskom súbore bola voľne odhadovaná. Tretím krokom bola metrická invariancia, kde boli ďalej obmedzené faktorové náboje (diskriminačné parametre) položiek v oboch skupinách na rovnaké hodnoty, zatiaľ čo rozptyl latentnej premennej v slovenskom súbore bol voľne odhadovaný. Konečne v poslednom, skalárnom modeli, bol uvoľnený aj priemer faktoru v slovenskom súbore, zatiaľ čo všetky parametre položiek boli obmedzené na rovnaké hodnoty. V prípade jednoparametrových raschovských modelov sme preskočili kroky jedna a dva (už úvodný model bol metricky invariálny, pretože faktorové náboje boli v oboch skupinách fixované na hodnotu 1).

Nakoniec sme jednotlivé modely porovnali pomocou pomeru vierohodností (LRT test) a pomocou indexov zhody s dátami. Na rozdiel od tradičnej analýzy invariance s pomocou konfirmačnej faktorovej analýzy nie sú k dispozícii odporúčania, ako veľký rozdiel v indexoch je už vecne významný; aj napriek tomu sme sa približne orientovali podľa bežných odporúčaní pre ordinálnu faktorovú analýzu (Putnick & Bornstein, 2016; Sass et al., 2014), pretože M_2^* štatistiku je možné interpretovať podobne ako χ^2 štatistiku vo faktorovej analýze. Pre porovnanie modelov sme využili tiež adjustované bayesovské kritérium (SABIC, sample-adjusted Bayesian Information Criterion).

Pre lepšie porozumenie prípadným príčinám noninvariance sme následne vykonali analýzu diferenciálneho fungovania položiek (DIF). Použili sme jedno- a dvojparametrový skalárny model a postupne odhadli pre každú položku separátne modely, v ktorom boli všetky parametre danej položky voľne odhadnuté v oboch skupinách. Ten sme následne prostredníctvom pomeru vierohodností a SABIC kritéria porovnali s pôvodným skalárne invariálnym modelom.

U položiek 21 a 24 sme pre účely analýzy invariance zlúčili odpoveďové kategórie 1 a 2 z dôvodu veľmi nízkej frekvencie správnych odpovedí, ktoré spôsobovali estimačné problémy.

Analytický kód je verejne dostupný na <https://doi.org/10.17605/OSF.IO/6TFV4>. Pretože je test v komerčnej distribúcii, dáta sú prístupné len na vyžiadanie u korešpondenčného autora.

VÝSLEDKY

Deskriptívne štatistiky

Priemerné skóre slovenského súboru a českého štandardizačného súboru je k dispozícii v tab. 5, v grafickej podobe na obr. 1. Je zrejme, že vo všetkých ročníkoch dosiahli slovenské deti v porovnaní s českými deťmi výrazne vyššieho skóre, pričom s rastúcim vekom sa tento rozdiel znižoval. Zatiaľ čo v treťom ročníku bol rozdiel 9,4 štandardného T-skóre s 95% intervalom spoľahlivosti $_{95\%}CI = [6,7-12,1]$, $t(151,5) = 6,83$, $p < 0,001$, a v štvrtom ročníku 8,5 s $_{95\%}CI = [5,4-11,7]$, $t(111,4) = 5,35$, $p < 0,001$, v piatom ročníku už bol rozdiel len 3,6 T-skóre s $_{95\%}CI = [0,9-6,3]$, $t(203,9) = 2,59$, $p = 0,010$.

Pre podrobný popis a exploračnú analýzu týchto rozdielov a pre kontrolu rozdielneho veku detí v českom a slovenskom súbore sme zostavili sériu lineárnych modelov, v ktorých bolo závislou premennou T-skóre a prediktory ročník, český vs. slovenský súbor, pohlavie a vek detí vrátane všetkých možných interakcií. Jediným štatisticky významným hlavným efektom bol opäť štát ($b = 9,6$ v treťom ročníku, $p < 0,001$), a ďalej bola signifikantná už len interakcia piateho ročníku a štátu ($b = -8,1$, $p = 0,022$); všetky ostatné $p > 0,15$. Z tab. 5 sa ďalej zdá, že variabilita hrubého skóre je u slovenských detí vyššia, to je však spôsobené zošíkmením skóre

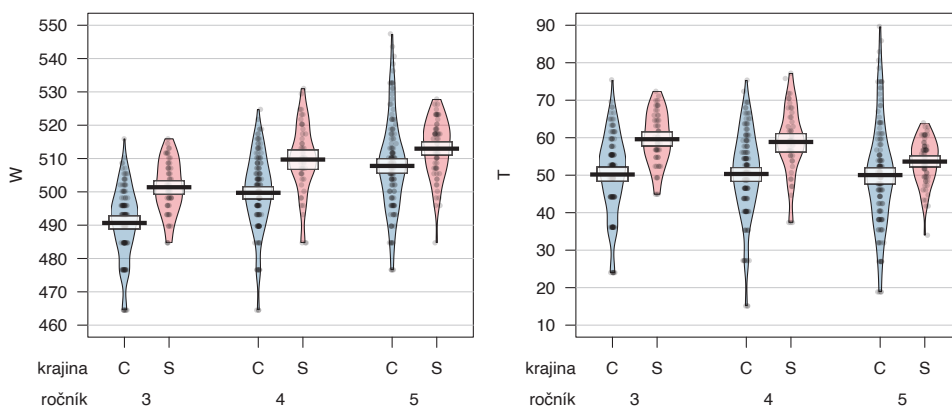
Tab. 5 Popisné štatistiky českého a slovenského výskumného súboru

	n	HS			W			T		
		M	SD	Sk	M	SD	Sk	M	SD	Sk
<i>český štandardizačný súbor</i>										
3. ročník	121	5,3	3,20	1,11	490,7	10,94	-0,54	50,2	10,94	-0,54
4. ročník	141	9,0	5,05	0,64	499,7	11,50	-0,64	50,3	11,49	-0,64
5. ročník	142	13,4	7,77	0,78	507,8	13,89	0,25	50,0	13,89	0,25
<i>slovenský súbor</i>										
3. ročník	54	9,3	4,04	0,33	501,4	7,98	-0,22	59,6	7,02	-0,22
4. ročník	51	14,3	6,13	0,16	509,7	10,56	-0,42	58,9	9,08	-0,42
5. ročník	64	16,2	5,45	-0,27	513,0	8,72	-0,64	53,6	6,06	-0,64

Poznámka. HS – hrubé skóre, W – Wskóre (IRT skóre), T – štandardné T-skóre s priemerom 50 a smerodajnou odchýlkou 10 v českom štandardizačnom súbore

a ich vyšším výkonom. Podstatnejší je rozdiel vo variabilite T-skóre, resp. v IRT odhadoch latentného rysu. Levenov test reziduí z lineárneho modelu potvrdzuje očividne menšiu variabilitu slovenských detí vo výkone, $F(1, 560) = 26,9, p < 0,001$. Pri pohľade na obrázok 1 sa zdá, že slovenský súbor je cenzurovaný, t.j. chýbajú deti s nižším výkonom, čo môže byť príčinou rozdielu v priemeroch medzi oboma súbormi.

Rozdiel vo výkone českých a slovenských detí sa teda nelíši naprieč tretňou a štvrtou triedou, v piatej je však menší. Vek detí nemá vplyv nad rámec ročníku a za pozorovaným rozdielom oboch súborov nestoja žiadne zo sledovaných demografických premenných. Ďalej sa nelíši ani výkon chlapcov a dievčat, a to ani v českom ani v slovenskom súbore. Aj po zohľadnení faktu, že slovenské deti boli staršie, zostáva rozdiel českého a slovenského súboru masívny a variabilita slovenského súboru je nižšia. Vzhľadom k existencii takto masívneho rozdielu sme v reakcii na recenzné posudky v priebehu recenzného procesu realizovali ešte dodatočnú exploračnú analýzu s využitím multilevel (hierarchickej) lineárnej regresie pomocou



Obr. 1 Rozdelenie výkonu českých (C) a slovenských (S) detí podľa ročníku na škále W-skóre (vľavo) a štandardizovaného T-skóre (vpravo)

R balíčkov lme4 a lmerTest (Bates et al., 2015; Kuznetsova et al., 2017). Ignorovanie hierarchickej štruktúry dát je napr. podľa Yarkoniho (2022) jednou z príčin krízy zo-všeobecniteľnosti v psychológii, pretože vedie k nadhodnoteniu štatistickej významnosti pozorovaných rozdielov a mohlo tak byť jednou z príčin štatisticky významného rozdielu medzi českými a slovenskými deťmi.

Závislou premennou bolo štandardné T skóre a definícia modelov bola zhodná ako v prípade lineárnych modelov vyššie (pevnými prediktormi teda bola krajina, ročník a ich vzájomné interakcie plus centrovany vek a pohlavie), doplnené o náhodné premenné v podobe školy a ročníku. Aj v tomto prípade vplyv krajiny, overený robustným Satterwhitovým t-testom, bol štatisticky významný, $B = 9,07$, aj keď signifikancia bola nižšia, $p = 0,040$; interval spoľahlivosti získaný profilovaním funkcie maximálnej vierohodnosti bol $95\% CI = [1,07-16,89]$. Rozdiel naprieč triedami bol relatívne malý, $SD = 2.09$ s $95\% CI = [0.00-3.36]$, $ICC = 0,032$, zato rozdiely naprieč školami boli zásadné, $SD = 6.22$ s $95\% CI = [3.86-8.96]$, $ICC = 0,284$, teda zhruba štvrtina variability medzi deťmi bola spôsobená rozdielmi medzi školami. Aj po zvážení hierarchickej štruktúry dát tak zostáva efekt krajiny štatisticky významný.

Reliabilita

Reliabilitu v českom aj slovenskom súbore sme odhadli dvoma spôsobmi. V prvom rade pomocou koeficientu alfa, ktorý je odhadom reliability súčtu všetkých položiek z hľadiska klasickej testovej teórie. Dôležitejším ukazovateľom je však reliabilita odhadu úrovne latentného rysu s pomocou Raschovho modelu; za týmto účelom sme využili jednoparametrový PCM model pre každý súbor zvlášť (viď ďalej). Z výsledkov v tab. 6 je zrejme, že reliabilita slovenskej formy testu je adekvátna a plne porovnateľná s reliabilitou českého súboru. Jediný významnejší rozdiel je v piatom ročníku, kde je test menej (avšak stále dostatočne) reliabilný v slovenskom súbore oproti českému; príčinou je zrejme nižšia variabilita a vysoký výkon slovenského súboru. Naopak v tretej a štvrtej triede je test v slovenskom súbore reliabilnejší, príčinou je opäť vyšší výkon slovenského súboru. Ďalej je tiež príčinou zdanlivo nízkej reliability v treťom ročníku v oboch súboroch vysoká obtiažnosť testu TIM³⁻⁵, ktorý preto diferencuje len v pásme vyššieho nadpriemeru.

Tab. 6 Reliabilita českého a slovenského súboru

	Český súbor					Slovenský súbor		
	táto štúdia			údaje z manuálu		táto štúdia		
	α	$r_{xx'}$	RMSE	α	$r_{xx'}$	α	$r_{xx'}$	RMSE
3. ročník	0,668	0,614	0,709	0,668	0,649	0,654	0,643	0,553
4. ročník	0,788	0,785	0,557	0,786	0,649	0,812	0,839	0,590
5. ročník	0,895	0,892	0,500	0,895	0,896	0,778	0,787	0,619
<i>M</i>	0,784	0,764	0,589	0,783	0,731	0,748	0,756	0,587

Poznámka. Údaje z manuálu českého súboru a údaje z tejto štúdie sa líšia najmä v prípade odhadu IRT reliability v dôsledku odlišných estimátorov Raschovho modelu. α – Cronbachova alfa; $r_{xx'}$ – odhad IRT reliability, v prípade tejto štúdie na základe 1PL modelu odhadnutého v balíčku mirt, v prípade údajov z manuálu s pomocou SW Winsteps; RMSE – priemerná chyba merania (root-mean squared residuum)

Tab. 7 Zhoda modelov s dátami v českom a slovenskom súbore zvlášť

	M_2^*	df	p	RMSEA	[90% CI]	SRMSR	TLI	porovnanie modelov				
								SABIC	$\Delta\chi^2$	Δdf	p	
CZ_1PL	466,9	288	<0,001	0,039	[0,033—0,046]	0,122	0,980	8975,1				
CZ_2PL	308,4	264	0,031	0,020	[0,007—0,029]	0,056	0,995	8888,6	154,3	24	<0,001	
SK_1PL	359,3	289	0,003	0,038	[0,023—0,050]	0,087	0,948	4571,9				
SK_2PL	322,0	265	0,009	0,036	[0,019—0,049]	0,073	0,954	4573,0	46,1	24	0,004	

Poznámka. 1PL – jednoparametrový Raschov model (PCM); 2PL – dvojparametrový model (GPCM). Vo všetkých prípadoch bol model definovaný ako jednodimenzionálny. Počty stupňov voľnosti českého a slovenského súboru nie sú totožné z dôvodu chýbajúce odpoved'ové kategórie.

Faktorová štruktúra českej a slovenskej verzie

V českom aj slovenskom súbore sme separátne odhadli jednoparametrový aj dvojparametrový IRT model a navzájom ich porovnali – výsledky sú v tab. 7. Aj napriek tomu, že Raschov model popísal dáta v oboch súboroch horšie než dvojparametrový model, vo všetkých prípadoch boli obidva modely veľmi dobré; a to okrem SRMSR

Tab. 8 Analýza invariancie

	M_2	df	p	RMSEA	[90% CI]	TLI	CFI	zrovnání modelov				
								SABIC	$\Delta\chi^2$	Δdf	p	
<i>Dvojparametrový GPCM model</i>												
konfi gurálna	634,7	530	0,001	0,019	[0,012—0,024]	0,992	0,993	13546,3				
prahová	658,8	540	<0,001	0,020	[0,014—0,025]	0,991	0,992	13541,5	27,0	10	0,003	
metrická	683,0	564	<0,001	0,019	[0,013—0,024]	0,991	0,992	13482,7	17,5	24	0,829	
skalárna	745,8	588	<0,001	0,022	[0,017—0,026]	0,989	0,989	13493,8	87,3	24	<0,001	
<i>Jednoparametrový PCM model</i>												
metrická	833,5	578	<0,001	0,028	[0,023—0,032]	0,982	0,982	13595,7				
skalárna	979,7	612	<0,001	0,032	[0,029—0,036]	0,976	0,974	13593,2	105,4	34	<0,001	

kritéria, ktoré môže byť nadhodnotené v dôsledku relatívne malých súborov a obzvlášť veľmi ťažkých položiek. SABIC kritérium jednoparametrového modelu v slovenskom súbore bolo dokonca lepšie než u modelu dvojparametrového.

Domnievame sa, že aj keď dvojparametrový model lepšie zodpovedá skutočnosti, jednoparametrový model je pre účely skórovania a vyhodnocovania testu viac než dostatočný.

Tab. 9 Výsledky analýzy diferenciálneho fungovania položiek

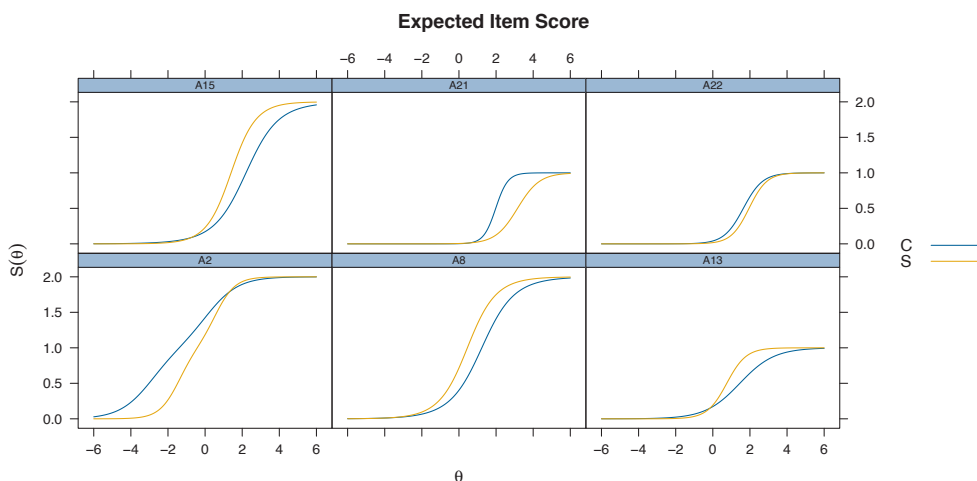
	SABIC	χ^2	df	p	p.adj
A1	5,30	1,056	2	0,590	1
A2**	-4,81	14,337	3	0,002	0,055
A3	0,54	5,812	2	0,055	1
A4	4,78	1,569	2	0,456	1
A5	7,86	1,672	3	0,643	1
A6	5,81	0,542	2	0,763	1
A7	6,28	0,076	2	0,962	1
A8***	-18,71	28,237	3	<0,001	<0,001
A9	7,55	1,982	3	0,576	1
A10	4,29	2,065	2	0,356	1
A11	7,76	1,767	3	0,622	1
A12	5,53	3,999	3	0,262	1
A13**	-6,44	12,796	2	0,002	0,038
A14	5,31	4,218	3	0,239	1
A15***	-8,71	18,235	3	<0,001	0,009
A16	2,12	7,405	3	0,060	1
A17	6,12	0,235	2	0,889	1
A18	5,76	0,597	2	0,742	1
A19	0,81	5,538	2	0,063	1
A20	2,98	3,372	2	0,185	1
A21*	-0,02	6,371	2	0,041	0,868
A22*	0,04	6,314	2	0,043	0,868
A23	2,02	7,506	3	0,057	1
A24	5,82	0,534	2	0,766	1
A25	5,68	0,671	2	0,715	1

Poznámka. p.adj – p-hodnota adjustovaná pomocou Holm-Bonferroniho korekcie; * $p < 0,05$;
 ** $p < 0,01$; *** $p < 0,001$ (bez adjustácie)

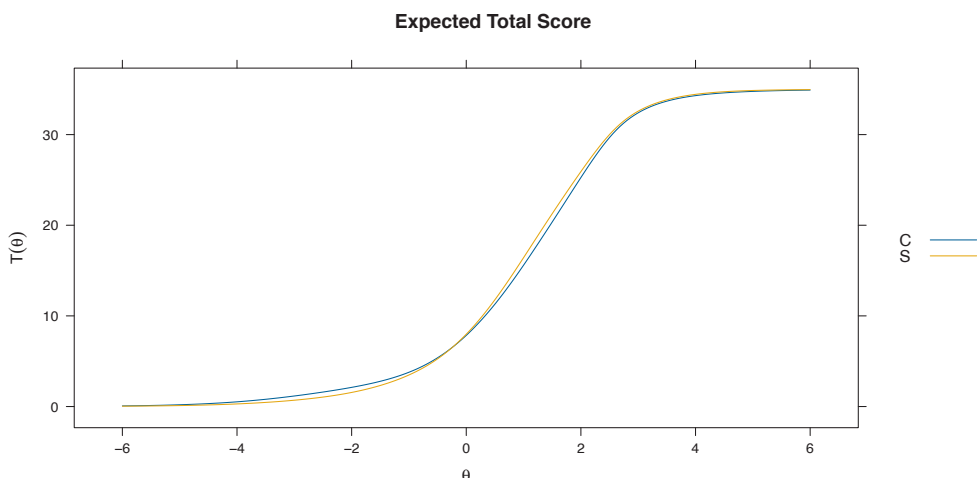
Analýza invariance

Invariancia českého a slovenského súboru na úrovni celého testu

Analýza invariance bola vykonaná separátne pre jednoparametrový a dvojparametrový model; výsledky sú v tab. 8. Aj keď viac obmedzené modely spravidla popisujú



Obr. 2 Grafy charakteristických funkcií položiek v českom (C; modrá) a slovenskom (S; ružová) súbore troch diferenciálne fungujúcich položiek z parciálne invariantného modelu. Na vodorovnej osi je miera latentného rysu, na zvislej osi očakávané skóre respondenta na danej položke.



Obr. 3 Graf charakteristickej funkcie testu pre český (C; modrá) a slovenský (S; ružová) súbor

jú dáta štatisticky významne horšie než menej invariantné modely, rozdiely indexov dobrej zhody sú zanedbateľné. V prípade dvojparametrového GPCM modelu sú dokonca informačné kritériá najlepšie pre metrickú invarianciu (a na druhom mieste je invariancia skalárna); v prípade jednoparametrového PCM modelu sú informačné kritériá porovnateľné (pričom skalárne-invariantný model je nepatrne lepší). Predložené výsledky považujeme za dôkaz skalárnej invariance testu TIM³⁻⁵ v českej a slovenskej populácii.

Analyza diferenciálneho fungovania položiek

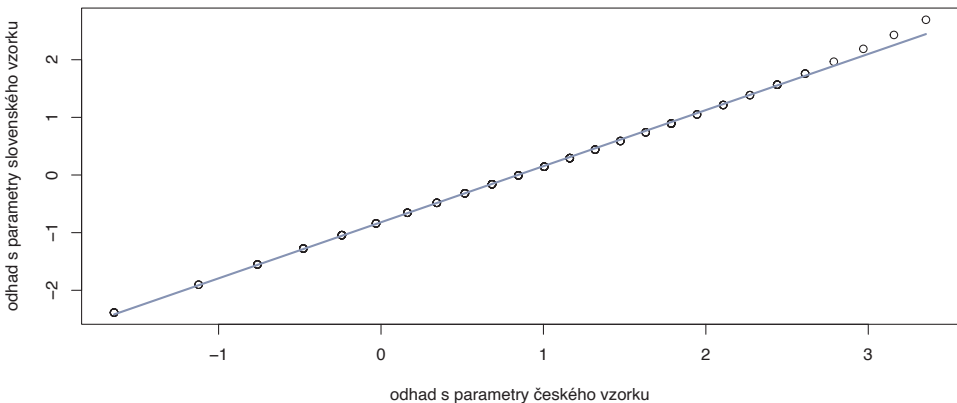
Tab. 9 obsahuje výsledky DIF analýzy. Vidíme, že parametre celkovo 6 položiek (24 % z celkového počtu položiek) sa štatisticky významne líšili v slovenskom a českom súbore; po Holm-Bonferroniho korekcii pre opakované testovanie ide len o 3 po-

ložky (12 %). To sa zdá byť pomerne vysoké číslo, obzvlášť pokiaľ zväžíme fakt, že model je celkovo skôr skalárne invariantný. Z týchto dôvodov sme odhadli nový parciálne-invariantný model, v ktorom sme u identifikovaných diferenciallyne fungujúcich položiek uvoľnili parametre naprieč oboma skupinami a preskúmali charakteristické funkcie položiek, viď obr. 2. Vplyv na skórovanie testu sa však zdá byť zanedbateľný, viď charakteristické funkcie testu pre obidve skupiny na obr. 3.

Vplyv non-invariance na skórovanie testu

Aby sme preskúmali vplyv prípadnej non-invariance na skórovanie testu v reálnej situácii, porovnali sme odhady IRT skóre slovenských respondentov získaných rôznym spôsobom. Prvý spôsob, ktorý slúžil ako benchmark, je originálne skórovanie použité v on-line vyhodnocovacej aplikácii českej verzie testu – výstupom je W-skóre. Druhý typ skóre bol odhadnutý metódou maximálnej vierohodnosti (ML) na základe jednoparametrového Raschova PCM modelu odhadnutého v balíčku mirt na súbore českých respondentov. Tretí spôsob bol totožný s druhým s tým rozdielom, že využil Raschov model odhadnutý priamo na slovenskom súbore.

Pearsonova korelácia originálne vytvoreného W-skóre so skóre na základe českého modelu bola $r = 0,9996$; W-skóre s odhadmi na základe slovenského modelu $r = 0,9987$; odhadov na základe českého a slovenského modelu $r = 0,9996$. Na obr. 4 vidíme, že vzťah odhadov faktorového skóre pri využití parametrov českého a slovenského súboru je prakticky perfektne lineárny a k nepatrným odlišnostiam dochádza až v pásme veľmi vysokého napriemeru (vzhľadom k rozdielnym výkonom českých a slovenských detí sa však líši referenčný, nulový bod škály).



Obr. 4 Porovnanie odhadu IRT skóre získaných pomocou modelu odhadnutého na českom a slovenskom súbore

Kvalitatívne zhodnotenie atypických odpovedí

Okrem správnych a typických chybných odpovedí sa opakovane objavilo aj niekoľko typov riešení, ktoré sa v českom súbore opakovane nevyskytli. V tejto kapitole budú stručne zhrnuté aj s informáciou o ich zastúpení v celom súbore detí a spôsobe bodovania týchto nejasných odpovedí. Presné znenie položiek nie je z dôvodu zachovania autorských práv vzťahujúcich sa k testu TIM³⁻⁵ uvedené.

Položkou, v ktorej sa netradičné odpovede vyskytovali najčastejšie, bola A13. Úloha je zameraná na priestorovú predstavivosť, pričom deti majú zakresliť pohľad na stavbu z kociek na základe uvedených pohľadov z iných smerov. Celkovo na túto položku od-

povedalo atypickým spôsobom 10 detí, čo predstavuje 7 % zo 142 detí, ktoré k položke uviedli nejakú odpoveď, a 5,9 % zo všetkých 169 detí. Z toho 4 deti (2,8 % zo 142) v odpovedi zakreslili objekt s hranou o dĺžke 4 kocky vedľa seba namiesto 3. Tento typ odpovede bol hodnotený ako nesprávny. Ďalej sa v tejto položke objavila tiež verzia, kedy deti uviedli v podstate správnu odpoveď, avšak prostrednú kocku označili ako chýbajúcu (3 deti, 2,1 %), prípadne hornú a spodnú hranu kociek prepojili iba jedinou kockou na strane (2 deti, 1,4 %), čo je z hľadiska priestorového videnia síce správna odpoveď, ktorá ale odporuje zákonom fyziky (stavba by okamžite spadla). Podľa inštrukcií v manuáli by ako správna mala byť hodnotená odpoveď, v ktorej je vyplnený spodný rad kociek a ktorýkoľvek jeden stĺpec, rozloženie ostatných kociek môže byť ľubovoľné. Všetky vyššie odpovede tieto podmienky spĺňajú, preto boli hodnotené ako správne. Žiadne z detí, ktoré uviedli rovnaký typ neobvyklej odpovede nepochádzali z rovnakej triedy, čo vylučuje možnosť, že by opakovanie sa danej odpovede bolo dané opisovaním.

Podobne atypické odpovede sa vyskytli v položke A3, kde je úlohou spočítať počet rezov potrebných pre rozdelenie tyče na štyri rovnaké časti. Celkovo tri deti (1,8 %) tyč rozdelili pozdĺžne, nie naprieč. Toto riešenie bolo uznané ako správne, pokiaľ okrem nákresu dieťa poskytlo aj adekvátnu číselnú odpoveď.

Grafické príklady všetkých chybných odpovedí s podrobnejším popisom sú k dispozícii v diplomovej práci Šintálovej (2022).

DISKUSIA

Zameraním tohto výskumu bola adaptácia českej testovej metódy TIM³⁻⁵ na Slovensko, pričom jeho súčasťou bol preklad testových materiálov, oslovenie slovenských škôl, zber dostatočného množstva dát a následná analýza so zameraním na preskúmanie prípadného odlišného fungovania slovenskej verzie testu, ktoré by mohlo narušiť ferovosť testu a tým pádom aj testovú validitu. Dôvodom k adaptácii tohto testu na Slovensko je jednak jazyková a kultúrna podobnosť Česka a Slovenska a zároveň nedostatok metód zameraných na meranie matematických schopností u detí v pásme nadpriemeru, o čom svedčia aj vyjadrenia oslovených psychológov z Centier pedagogicko-psychologického poradenstva a prevencie na Slovensku. Predpokladom úspešnej adaptácie testu je zohľadniť a pokiaľ možno eliminovať možné zdroje invalidity, ktorými môžu byť kultúrne alebo jazykové rozdiely, technické a metodologické nedostatky výskumu, prípadne nevhodná interpretácia výsledkov (Hambleton, 2005).

Rozdiely vo výkone slovenských a českých detí

Aj napriek očakávaniam, že by sa slovenské a české deti v priemerne dosiahnutom skóre z celého testu nemali líšiť, slovenské deti dosiahli signifikantne vyššieho výsledku, ktorý sa prejavil na úrovni všetkých ročníkov. Rozdiel v 5. ročníku bol však menší ako pri 3. a 4. ročníku. Vzhľadom k tomu, že v medzinárodných testovaniach PISA (Schleicher, 2019) a TIMSS (Mullis et al., 2020) dosahovali dlhodobo české deti mierne lepších až totožných, rozhodne však nie horších výsledkov, je preto nepravdepodobné, že by slovenské deti mali mať skutočne lepšie matematické schopnosti. Prípadný rozdiel medzi oboma populáciami by sa tiež mal prejavovať obdobne naprieč všetkými ročníkmi, nemal by byť v piatej triede nižší. Vysvetlenie tohto rozdielu je preto vhodnejšie hľadať skôr v metodologických odlišnostiach medzi českým a slovenským výskumom – výber ani jedného výskumného súboru nebol náhodný (ani napr. kvótny) a ich reprezentativita je preto diskutabilná. Inými slovami, považujeme za vysoko nepravdepodobné, že pozorovaný rozdiel v priemerných výkonoch oboch súborov reflektuje skutočné rozdiely českej a slovenskej populácie.

Prvým z možných vysvetlení by mohol byť odlišný výber škôl. Aj napriek tomu, že zámerne neboli (z dôvodu možnej odlišnosti od zvyšku Slovenska) oslovované školy z Bratislavy, zahrnuté školy pochádzali z väčších slovenských miest (Žilina, Trnava a Trenčín), ktoré majú všetky viac ako 55 tisíc obyvateľov. V českom výskume bolo zahrnutých 13 škôl, pričom 11 z nich pochádzalo z obce s počtom obyvateľov nižším ako 10 tisíc; značná časť súboru pochádzala napr. z Ústeckého kraja, v ktorom žiaci dlhodobo vykazujú nižšie študijné výsledky (Lebeda et al., 2022).

Zároveň však neexistuje presvedčivý dôkaz o tom, že by veľkosť miesta bydliska mala výrazný vplyv na matematické schopnosti detí, je preto nepravdepodobné, že by tento faktor mohol byť vysvetlením zisteného rozdielu. Hypotézu o rozdieloch škôl čiastočne podporuje náš dodatočný exploračný model hierarchickej regresnej analýzy. Rozdiely medzi školami boli značné a vysvetlili zhruba štvrtinu variability detí, avšak hypotézu nie je možné podporiť jednoznačne, pomocou rôznych štatistických postupov sme došli k rôznym záverom. Nie je však možné vylúčiť, že skutočný rozdiel oboch krajín je výrazne menší, 95% interval spoľahlivosti sa pohyboval v rozmedzí 1,07–16,89 štandardného T-skóre.

Významným rozdielom, ktorý by naopak mohol do veľkej miery vysvetliť lepšie výsledky slovenských detí, je zbieranie informovaných súhlasov, pričom spôsob akým tento fakt mohol ovplyvniť výsledky je nasledovný. K zapojeniu dieťaťa do výskumu bolo potrebné získať podpísaný súhlas od jeho rodiča alebo zákonného zástupcu. Súhlasy sa rozдали deťom, ktoré mali rodičov oboznámiť s tým, že majú možnosť zúčastniť sa tohto výskumu a pokiaľ rodič aj dieťa súhlasili, dieťa prinieslo podpísaný informovaný súhlas naspäť do školy a zúčastnilo sa zberu dát. Avšak, k narušeniu tohto procesu mohlo dôjsť na úrovni dieťaťa, ako aj jeho rodiča. Dieťa, ktoré o svojich matematických schopnostiach nemá vysokú mienku, prípadne nerado rieši matematické úlohy, sa po prečítaní informácie, že účasť je dobrovoľná mohlo rozhodnúť, že o výskum nemá záujem. K rovnakej reakcii a rozhodnutiu súhlas neudelit' mohlo dôjsť aj zo strany rodičov, ktorí sú si vedomí podpriemerných matematických schopností svojho dieťaťa. Pokiaľ dieťa ani rodič neočakávajú dobré výsledky, motivácia k zapojeniu sa do výskumu je nižšia, a tým pádom z výskumu systematicky môžu vypadávať deti s nižšími schopnosťami. Keďže návratnosť súhlasov bola priemerne okolo 50 %, vplyv takéhoto systematického úbytku sa mohol prejaviť vo vyššom priemernom skóre zapojených detí. Avšak vzhľadom k tomu, že cieľom tejto práce nebola štandardizácia a vytváranie slovenských noriem, nereprezentatívnosť dát by nemala spôsobovať závažný problém. Pre porovnanie, podľa autorov testu (osobná komunikácia) bola návratnosť dát v českej štandardizácii blízko 100 %. Domnievame sa preto, že rozdiel v priemernom výkone českých a slovenských žiakov mohol byť do značnej miery spôsobený práve rozdielmi v tvorbe výskumného súboru. Túto domnienku podporuje aj porovnanie tvaru distribúcie štandardného skóre českého a slovenského súboru. Výkon slovenských žiakov bol menej variabilný a pôsobil cenzo- rovaným dojmom, akoby ľavá časť distribúcie chýbala, čo odpovedá našej hypotéze.

Ďalším potenciálnym faktorom ovplyvňujúcim dosiahnuté výsledky slovenských detí by mohol byť fakt, že asi tretina dát bola zbieraná školskou psychologičkou, a nie výskumníkom. Pri porovnaní škôl sa ukázalo, že škola, na ktorej boli dáta zbierané výhradne školskou psychologičkou, skutočne dosiahla o niečo vyšších priemerných výsledkov, zároveň sa však tento rozdiel nepreukázal byť štatisticky významný. Táto psychologička pôsobila ešte aj na druhej škole, v ktorej zozbierala asi 65 % dát, táto škola sa však dosiahnutými výsledkami od ostatných nelíšila, neexistuje preto stabilný dôkaz o tom, že by deti, ktorým bol test administrovaný psychologičkou, dosahovali lepších výsledkov.

Za zmienku stojí tiež rozdiel v priemernom veku žiakov medzi českým a slovenským súborom. Slovenské deti boli priemerne asi o 2 mesiace na ročník staršie, tento rozdiel je vysvetliteľný o trochu neskorším zberom dát oproti českému súboru. Táto drobná veková výhoda však potenciálne mohla prispieť k lepším výkonom slovenských detí, pričom v súlade s výsledkami by pri nižších ročníkoch bol tento rozdiel výraznejší.

Dalším rozdielom medzi slovenským a českým výskumným súborom je skutočnosť, že medzi zbermi dát v Českom a slovenskom výskume prebehla vlna pandémie COVID-19, ktorá vo výraznej miere ovplyvnila spôsob a potenciálne aj kvalitu vzdelávania na Slovensku. Vplyv, ktorý by sme pravdepodobne na výkon detí očakávali, by však mal byť skôr negatívny pre slovenské deti, ktoré boli merané po pandémii. Možnou hypotézou, vysvetľujúcou túto nezrovnalosť je, že dočasné zatváranie škôl a presun do online výuky z dlhodobého hľadiska nemalo tak negatívny efekt na výsledky detí, ako bolo pôvodne očakávané, obzvlášť pri deťoch s celkovo lepšími študijnými výsledkami (Schult et al., 2022). Ďalšou možnosťou je, že keďže TIM³⁻⁵ bol cielene konštruovaný so zámerom zachytávať vplyv naučených znalostí a schopností čo najmenej, mierne zníženie kvality školských znalostí u detí by sa v ňom nemuselo prejavovať. Z uvedených dôvodov teda považujeme za vysoko nepravdepodobné, že pozorovaný rozdiel českého a slovenského súboru reflektuje skutočný rozdiel českej a slovenskej populácie. Príčinami rozdielu sú skôr systematická výberová chyba v jednom alebo oboch súboroch a nízka miera návratnosti informovaných súhlasov v slovenskom súbore. Menší až zanedbateľný rozdiel mohol mať rozdielny priemerný vek oboch súborov, ďalšie zvažované príčiny (osoba administrátora, pandémie covidu-19, prípadne non-invariancia testu – viď nižšie) sú nepravdepodobné. Je možné, že české normy boli zostavené z podpriemerných detí a môžu teda byť príliš „mäkké“. Naše výsledky preto považujeme za vysoko relevantné aj pre českú odbornú komunitu. Považujeme za vhodné upozorniť českého užívateľa testu na toto riziko a odporučiť preto zvýšenú opatrnosť pri interpretácii testových výsledkov, prinajmenšom do doby, než bude reprezentatívnosť českých noriem overená ďalšou nezávislou štúdiou. Na základe súčasných zistení je možné odporučiť, aby aj výsledky českých detí boli interpretované skôr konzervatívne.

Psychometrické parametre slovenskej adaptácie

Slovenský súbor vykázal primerané hodnoty reliability, porovnateľné s českým manuálom, alebo dokonca lepšie. Rozdiely však môžu súvisieť s tým, že test lepšie diferencuje v pásme nadpriemeru a slovenské deti mali v priemere vyšší výkon. Test mal tiež podľa očakávaní jednodimenzionálnu štruktúru podobne ako v českom súbore (Cígler et al., 2017a).

Rozdiely vo fungovaní testu českých a slovenských detí (invariancia merania)

Kľúčovým zistením pre adaptáciu testu do slovenského prostredia je fakt, že sa parametre položiek v českom a slovenskom súbore prakticky nelíšia. Test sa ukázal ako skalárne invariantný, čo znamená, že meraný latentný rys je rovnaký v oboch populáciách (má rovnakú mierku aj referenčný bod) a skóre v oboch súboroch a populáciách sú plne porovnateľné. Tento záver je pritom pomerne prekvapivý. V rámci českého súboru totiž položky vykazovali drobné rozdiely vo fungovaní podľa veku a absolútnych schopností respondentov (Cígler et al., 2017a). Podobnú non-invarianciu sme preto očakávali pri porovnaní českej a slovenskej verzie, pretože slovenské deti podali priemerne vyšší výkon.

Podrobnejšia analýza diferenciálneho fungovania položiek naznačila drobné rozdiely pri šiestich položkách, po kontrole štatistickej významnosti na opakované testovanie sa signifikantne líšili parametre len pri troch položkách (A8, A13 a A15). Všetky boli jednoduchšie pre slovenské deti (a to aj po kontrole celkovo vyššieho výkonu slovenského súboru), položka A13 by navyše aj lepšie diskriminovala u slovenských detí (pokiaľ by bol použitý viacparametrový IRT model). Zatiaľ čo pri položkách A8 a A15 nemáme hypotézu vysvetľujúcu uvedené rozdiely, pri položke A13 by mohla hrať rolu systematická odlišnosť v štýle práce detí, ktorá sa prejavila aj vyšším podielom atypických odpovedí, ktoré sa vôbec nevyskytli v priebehu českej štandardizačnej štúdie. Tento efekt sa však neopakoval pri položke A3, ktorá mala taktiež vysoký podiel atypických odpovedí, ale diferenciálne fungovanie nevykazovala.

Záver: Použitelnosť českých noriem testu TIM³⁻⁵ pre slovenské deti

Psychometrické parametre testu TIM³⁻⁵ na slovenskom pilotnom súbore boli veľmi dobré a parametre položiek sa od českej populácie líšili len minimálne. Je teda možné konštatovať, že test meria rovnaký latentný rys – matematické schopnosti – v českej aj slovenskej populácii žiakov 3.–5. ročníkov základných škôl, testové skóre je priamo porovnateľné a test je možné použiť pre meranie matematických schopností slovenských detí.

Zásadnou prekážkou pre využitie testu pre slovenské deti je však rozdiel v priemernom výkone. Použitie českých noriem pre slovenských respondentov totiž predpokladá ekvivalentné rozloženie výkonu v oboch populáciách. Hoci je tento predpoklad teoreticky samozrejмый, nie je podporený dátami. Slovenské deti naopak podali v priemere takmer o jednu smerodajnú odchýlku vyšší výkon, pričom príčina tohto rozdielu nie je známa. Vzhľadom k tomu, že takto veľký rozdiel v schopnostiach českej a slovenskej populácie je nepravdepodobný, ponúkame alternatívne vysvetlenie založené najmä na spôsobe výberu výskumného súboru (podiel návratnosti informovaných súhlasov) a konkrétnej výberovej populácii (veľkosť sídla a podobne). Napriek tomu rozdiel limituje využitie noriem – užívateľ testu si musí byť vedomý, že prípadného slovenského respondenta porovnáva s českou štandardizačnou vzorkou, nie slovenskou, a štandardné skóre tak môže nadhodnotiť porovnanie so slovenskou populáciou, a to až o jednu smerodajnú odchýlku. Bezmyšlienkovité a rutinné využitie slovenských noriem je preto nevhodné.

Zistenie je relevantné aj pre českého užívateľa, pretože bez ohľadu na kvalitu ostatných psychometrických parametrov testu môže naznačovať jemný posun českých noriem v smere väčšej miernosti. Považujeme za vhodné realizovať ďalšie empirické štúdie a overiť prinajmenšom kvalitu českých noriem.

Na základe našich výskumných výsledkov ponúkame test k použitiu slovenskými psychológmi a pedagógmi. Odporúčame však opatrnú interpretáciu štandardného skóre (T-skóre a percentil); pre slovenské deti môžu byť dostupné normy príliš „mäkké“. Pre porovnanie slovenských detí navzájom je však test vhodný a pretože na Slovensku neexistuje iná alternatíva, domnievame sa, že je možné ho s vedomím vyššie uvedených rizík využívať aj v bežnej pedagogicko-psychologickej praxi. V takom prípade môže byť vhodné zvoliť prísnejšiu hodnotu kritického skóre pre identifikáciu nadania než je bežné štandardné T skóre 70.

LITERATÚRA

Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2005). *Test struktury inteligence IST 2000 R*. Testcentrum.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*

- re, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Cai, L. & Monro, S. (2014). *A new statistic for evaluating item response theory models for ordinal data*. National Center for Research on Evaluation, Standards, & Student Testing. Technical Report.
- Cígler, H., Jabůrek, M., Straka, O., & Portešová, Š. (2017a). *Psychometrická analýza TIM³⁻⁵ - Testu pro identifikaci nadaných žáků v matematice pro 3.-5. třídu*. Masarykova univerzita.
- Cígler, H., Jabůrek, M., Straka, O., & Portešová, Š. (2017b). *Test pro identifikaci nadaných žáků v matematice pro 3.-5. třídu*. Masarykova univerzita.
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1–29. <https://doi.org/10.18637/jss.v048.i06>
- Dočkal, V. (2017). *I-S-T 2000 R: Test štruktúry inteligencie*. Hogrefe - Testcentrum.
- Dočkal, V. (2000). *SOCAG*. VÚDPaP.
- Fay, E., Trost, G., & Gittler, G. (2001). *Intelligenz-Struktur-Analyse (ISA)*. Swets Test Services.
- Ferjenčík, J., & Hromý, J. (1989). *Ravenove progresívne matice*. Psychodiagnostika.
- Kaufman, A. S., & Kaufman, N. L. (1999). *Kaufmanova hodnotaica batéria pre deti K-ABC*. Psychodiagnostika.
- Košč, L. (2002). *Kohsove kocky*. Psychodiagnostické a didaktické testy.
- Košč, L. (1987). *Test matematických schopností: Kalkúlie III, Číselný trojuholník, Rey-Osterriethova komplexná figura*. Psychodiagnostika.
- Krčová, V., Portešová, Š., & Cígler, H. (2017). Reyova-Osterriethova komplexná figura ve vztahu k rozumu nadání a úrovni matematických schopností u dětí z třetich až pátých ročníků základních škol. *Československá psychologie*, 61(1), 3–15.
- Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lebeda, T., Lysek, J., Marek, D., Brusenbauch Meislová, M., Soukup, M., Zymová, K., Zapletalová, M., Basl, J., Novosák, J., Zatloukal, T., Macková, B., Janega, J., Spitzerová, M., Hájek, O. & Daniel, S. (2022). *Zajímavosti českého vzdělávání: Výbrané faktory ovlivňující podmínky, průběh a výsledky vzdělávání pohledem pěti let sekundárních analýz*. Česká školní inspekce. Získané z https://www.csicr.cz/CSICR/media/Prilohy/2023_p%C5%99%C3%ADlohy/Dokumenty/Sekundarni_analyza_Zajimavosti_ceskeho_vzdelavani_finalPT.pdf
- Metodické pokyny na zaraďovanie detí do špeciálnych výchovno-vzdelávacích programov pre intelektovo nadaných žiakov č. CD-2005-19376/26377-1:091, schválené Ministerstvom školstva SR 25. 8. 2005. Získané z https://www.cppap.sk/data/2005_19376_aktual.pdf
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International results in mathematics and science*. Získané z <https://timssandpirls.bc.edu/timss2019/international-results/>
- Novák, J. (2001). *Barevná kalkúlie*. Psychodiagnostika.
- Perleth, C., Schatz, T., & Mönks, F. J. (2000). Early indicators of high ability. In K. A. Heller, F. J. Mönks, R. Subotnik, & R. J. Sternberg (Eds.), *International handbook for giftedness and talent* (2nd edition, pp. 297–316). Elsevier.
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Ruef, M., Furman, A., & Muñoz-Sandoval, A. (2003). *Woodcock-Johnson. Medzinárodná edícia*. The Woodcock-Muñoz Foundation.
- Ruef, M., Furman, A. & Muñoz-Sandoval, A. (2010). *Woodcock-Johnson Czech International Edition II*. The Woodcock-Muñoz Foundation.
- Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(2), 167–180. <https://doi.org/10.1080/10705511.2014.882658>
- Schleicher, A. (2019). *PISA 2018: Insights and interpretations*. OECD.
- Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues*. The Guilford Press.
- Schrank, F. A., McGrew, K. S., Mather, N., Wendling, B. J., LaForte, E. M. (2014). *Woodcock-Johnson IV: Tests of Achievement*. Riverside Publishing Company.
- Schult, J., Mahler, N., Fauth, B., & Lindner, M. A. (2022). Did students learn less during the COVID-19 pandemic? Reading and mathematics competencies before and after the first pandemic wave. *School Effectiveness and School Improvement*, 33(4), 544–563.
- Šintálová, M. (2022). *Adaptace testu TIM3-5 do slovenského prostredia* [Diplomová práca, Masarykova univerzita]. <https://is.muni.cz/th/ylkz1/>

- Terman, L. M., & Merrill, M. A. (1972). *Stanford-Binetova inteligenčná škála*. Psychodiagnostika.
- Thorndike, R., I., & Hagen, E. (1998). *Test kognitívnych schopností – TKS*. Psychodiagnostika.
- Urban, K. K., Jellen, H. G., & Kováč, T. (2002). *Urbanov Figurálny test tvorivého myslenia – TSD-Z*. Psychodiagnostika.
- Von Aster, M. G., & Weinholdová, M. (2008). *ZAREKI: Neuropsychologická batéria testov na spracovávanie čísiel a počítanie u detí*. Psychodiagnostika.
- Vonkomer, J., & Miglierini, B. (1992). *Názorové rady*. Psychodiagnostika.
- Wechsler, D., Dočkal, V., Kretová, E., Kundrátová, B., Sedláčková, B., Tesař, M. (2006). *WISC-III – Wechslerova inteligenčná škála pre deti*. Testcentrum – Hogrefe.
- Wechsler, D., Krejčířová, D., Boschek, P., & Dan, J. (2002). *WISC-III – Wechslerova inteligenční škála pro děti*. Testcentrum Praha.