

Test prostorových schopností Lin a vesmírná navigace – technický manuál – verze 1.0 (19.12.2024)

Aktuální verzi citujte jako: Černický I., Jabůrek, M., Palíšek, P., Straka, O. & Portešová, Š. (2024). Test prostorových schopností Lin a vesmírná navigace - technický manuál, verze 1.0. Invenio. Dostupné z www.invenio.muni.cz.

Obsah

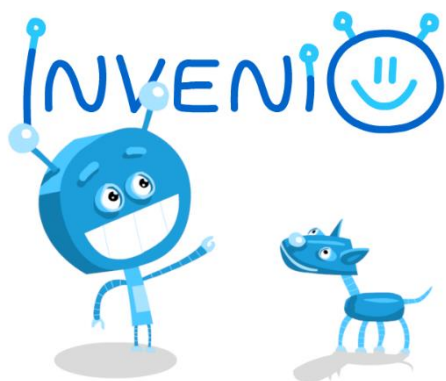
Psychodiagnostický systém Invenio.....	2
Autoři	2
Účel a možnosti použití.....	3
Obecná specifikace testů	3
Podoba výstupních zpráv	4
Lin a vesmírná navigace.....	4
Východiska a zaměření	5
Teoretická východiska.....	5
Existující testové principy a podobně zaměřené testy	6
Časový limit	7
Popis a struktura	8
Herní příběh.....	8
Struktura a jednotlivé prvky testu	8
Vývoj testu a odpověďové procesy	10
První pilotní studie	11
Druhá pilotní studie	11
Standardizační studie.....	11
Vzorek	11
CTT položková analýza	12
IRT položková analýza.....	13
Invariance Raschova modelu	18
Post-standardizační (finální) podoba testu	18
Vzorek	19
Reliabilita	19
Normy.....	19
Férovost.....	20
Validita	21
Poděkování	22
Literatura	22

Cílem tohoto manuálu je nejdříve stručně představit diagnostický systém Invenio a následně detailněji informovat čtenáře o testu *Lin a vesmírná navigace* (zkráceně *Lin*), tedy popsat jeho východiska, vývoj, strukturu a předložit důkazy o jeho kvalitě. Manuál představuje test *Lin* jako moderní a transparentně vyvíjenou metodu odpovídající mezinárodně uznávaným nárokům na nástroje daného typu. Při vývoji testu byly respektovány Standardy pro pedagogické a psychologické testování (AERA, APA, & NCME, 2014)¹.

Psychodiagnostický systém Invenio

Invenio je online psychodiagnostický systém určený ke screeningu profilu kognitivních schopností žáků 1.–6. tříd. Vychází z empiricky podložené psychologické teorie C-H-C (Cattell–Horn–Carroll, např. Schneider & McGrew, 2018), dle níž se inteligence skládá z řady dílčích schopností jako je např. *Fluidní inteligence*, *Krátkodobá pracovní paměť*, *Vizuální zpracování* a dalších. Každý test systému je přitom zaměřen na jednu z těchto schopností. Při testování je současně využíván moderní přístup tzv. *game-based assessmentu*, kdy jsou jednotlivé testy pojaty jako jednoduché počítačové hry, do kterých jsou integrovány herní prvky jako např. příběh, komiksová grafika, zvuky, sbírání bodů apod. Díky tomu obvykle dochází ke zvyšování motivace, navození pozitivního prožitku při řešení úloh a redukci případných obav u testovaných dětí (Hawkins et al., 2013; Kiili & Ketamo, 2017).

Níže popsaný test s názvem *Lin a vesmírná navigace* (zkráceně *Lin*) zaměřený na *Prostorové schopnosti* (konkrétně schopnost *Vizualizace*) je jednou z řady diagnostických her, které jsou součástí systému Invenio. V aktuální podobě je určen žákům 1.-6. tříd základních škol.



Obrázek 1. Logo a hlavní postavy systému Invenio.

Autoři

Systém Invenio vyvinuli odborníci z Katedry psychologie Fakulty sociálních studií Masarykovy univerzity v Brně. Více informací o systému, jednotlivých členech týmu a dalších testech naleznete na www.invenio.muni.cz.

Odborný tým můžete kontaktovat na info@invenio.muni.cz.

¹ Tyto standardy, které jsou pravidelně aktualizovány, jsou výsledkem spolupráce tří organizací - American Educational Research Association, American Psychological Association a National Council on Measurement in Education.

Účel a možnosti použití

Systém Invenio je screeningovým nástrojem, který je zaměřen na jednotlivé úzké schopnosti v rámci inteligenční teorie C-H-C. Může být proto využit všude, kde je účelné získat orientační představu o jednotlivých kognitivních schopnostech žáka.

Získané výsledky mohou pomoci dětem, pracovníkům školy (pedagogům, školním psychologům či speciálním pedagogům) či rodičům při individualizaci vzdělávacího přístupu, k zamyšlení nad dalším rozvojem měřených schopností, při zvažování dalšího profesního směřování (volba školy či vzdělávacího kurzu) apod. Hlavním účelem systému je přitom vyhledávání nadaných žáků základních škol. Systém je vyvinut pro plošné testování ve školách, a z tohoto důvodu je koncipován tak, aby se ho mohli účastnit i žáci, u nichž lze očekávat spíše průměrný nebo podprůměrný výkon. Testy proto obsahují položky různé obtížnosti, jejich normy jsou vytvořeny pro běžnou populaci žáků, a celý systém využívá herní principy, které obvykle snižují testovou úzkost.

Samotné zadávání testu je velmi snadné, protože žáci postupují dle instrukcí, které jsou součástí hry. Úkolem zadávajícího je především zajistit standardní podmínky. Testy tak může zadávat jakákoliv zaškolená osoba (např. pracovník školy). V aktuální podobě není možné testy zadávat v domácím prostředí žáka, aby nebyla ohrožena standardnost a validita testování.

Testování nemohou být žáci se závažnými poruchami zraku, sluchu a jemné motoriky, děti se závažným mentálním znevýhodněním či žáci, kteří nerozumí češtině, protože by mohla být ohrožena validita získaných výsledků. Testy je možné administrovat i žákům méně zběhlým v užívání počítače. Je pouze nutné se ujistit, že jsou děti (především z prvního ročníku) schopny samostatně ovládat počítačovou myš.

Obecná specifikace testů

Jednotlivé testy jsou administrovány prostřednictvím počítače. Jsou spouštěny online v rámci webového prohlížeče. Je možné je zadávat skupinově, nicméně každý žák pracuje individuálně na svém počítači. Hry ovládá pomocí myši a klávesnice, zvukové podněty přijímá prostřednictvím sluchátek.

Celý diagnostický systém je z pohledu hráče situován do vesmírného prostředí. Obecným úkolem je plnit různorodé úkoly na jednotlivých vesmírných planetách. Každá planeta přitom představuje psychodiagnostický test zaměřený na jednu z úzkých schopností teorie C-H-C. V základním hráčském menu jsou zobrazeny všechny planety (testy), které má hráč v rámci daného testového sezení prozkoumat. Již ukončené testy jsou graficky odlišeny od těch, které bude hráč teprve řešit. Tím je dosahováno větší názornosti jak pro hráče, tak pro administrátora.

V rámci herního příběhu vystupují dvě postavy – robot P.U.K (První Univerzální Kosmonaut) a jeho psí robotický kamarád Vincent van Dog. Ti provádí hráče testem, předávají mu instrukce k řešení jednotlivých testových položek a upozorňují jej na případné chyby. Větší srozumitelnosti je dosahováno jednoduchými animacemi (ukazování na některý prvek, přesun prvku po obrazovce apod.).

Kromě řešení jednotlivých úloh může hráč využít dvě tlačítka (viz Obrázek 2) – pauzu, která slouží k dočasnému zastavení hry, a vypnutí zvuků, kterým se dají eliminovat doprovodné herní zvuky, jako je hudba, klikání apod. Zvuky, které jsou zásadní pro ovládání hry a řešení úloh, jsou slyšet stále a není možné je vypnout.



Obrázek 2. Ovládací prvky totožné pro všechny testy systému Invenio – pauza a vypnutí doprovodných zvuků.

Podoba výstupních zpráv

Po testování systém Invenio vygeneruje dva dokumenty s výsledky testovaného dítěte – *Zprávu pro rodiče* a *Odbornou zprávu*.

Cílem *Zprávy pro rodiče* je poskytnout zákonným zástupcům dítěte základní informace o měřených schopnostech, výsledcích, jejich interpretaci a na ně navázaných doporučeních. Výsledky jsou uváděné ve formě zařazení do jedné ze čtyř kategorií, které jsou odvozené od bodové hodnoty standardního skóre (s průměrem 100 a směrodatnou odchylkou 15). Tyto kategorie slouží uživatelům pro snadnější interpretaci výsledků. Jednotlivé kategorie s příslušným rozmezím standardních skóre jsou k dispozici v následující Tabulce 1.

Tabulka 1. Kategorie a jejich vztah ke standardnímu skóru.

Kategorie	Rozmezí standardního skóru
Snížená schopnost	85 a méně*
Průměrná schopnost	85-114
Rozvinutá schopnost	115-129
Mimořádně rozvinutá schopnost	130 a více

Pozn.: *Kategorie snížená schopnost je uvedena ve chvíli, kdy bodový odhad standardního skóre činí 85 a méně bodů a zároveň horní mez intervalu spolehlivosti činí 90 a méně bodů.

Dokument označený jako *Odborná zpráva* je odborným výstupem z psychologické diagnostiky, který je uživatelům poskytován pro účely případných navazujících vyšetření a odborných konzultací (např. v pedagogicko-psychologické poradně). Obsahuje technické informace o použitých diagnostických hrách, měřených schopnostech a výsledky dítěte ve formě standardních skóre. Ty jsou uváděny v podobě 95% intervalu spolehlivosti, který je odvozen od chyby měření testu.

Lin a vesmírná navigace

Test *Lin a vesmírná navigace* (krátce *Lin*) je jednou z diagnostických her systému Invenio. Byl navržen pro měření *Vizuálního zpracování* (*Gv*), a to konkrétně se zaměřením na úzkou schopnost *Vizualizace* (*Vz*) a částečně také *Vizuální paměť* (*MV*). V současné podobě je test normován pro 1. – 7. třídu základních škol. Jeho hlavním účelem je screening kognitivních schopností a identifikace dětí, které by mohly těžit z rozvojových programů pro mimořádně nadané v této oblasti. V kombinaci s ostatními metodami diagnostického systému Invenio nabízí také podrobnou představu o profilu kognitivních schopností žáků, jejich silných a slabých stránkách.

Narozdíl od předchozích testů *Vizuálního zpracování* pracuje metoda *Lin* s animovanými podněty. Výsledky tedy odrážejí tzv. dynamické prostorové schopnosti, které jsou relevantní v mnoha oblastech života, zejména v prostředí informačních technologií. Test byl vyvíjen pomocí moderních psychometrických postupů (IRT). V manuálu jsou pro srovnání k nalezení i statistiky z klasické testové teorie (CTT).

Východiska a zaměření

Teoretická východiska

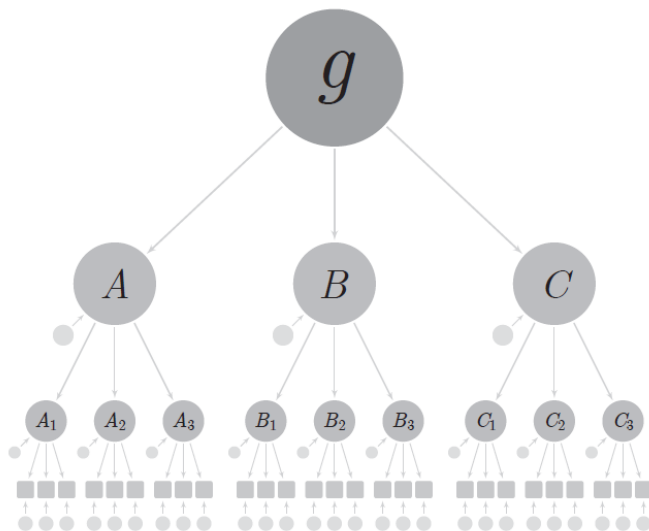
Test *Lin* vychází z klasifikace kognitivních schopností C-H-C teorie, která je popsána níže.

Teorie inteligence C-H-C

Cattell–Horn–Carrollova teorie inteligence (C-H-C) je komplexní taxonomií kognitivních schopností, která v sobě integruje několik předešlých teorií inteligence (např. Schneider & McGrew, 2018). Hlavními stavebními kameny jsou Cattell-Hornova *Gf-Gc* teorie (např. Horn & Noll, 1997) a Carrollova teorie tří vrstev (v angl. *The Three-Stratum Theory of Cognitive Abilities*; Carroll, 1993). Jde o psychometrické přístupy ke studiu inteligence, které využívají primárně faktorově-analytický přístup pro vysvětlení různé míry korelace mezi kognitivními schopnostmi.

Aktuální C-H-C teorie (Schneider & McGrew, 2018) pracuje s hierarchií tří vrstev kognitivních schopností (ilustrováno na obrázku 3 níže). Na vrcholu stojí obecná inteligence (historicky označována jako *g* faktor), která ovlivňuje nejširší životní oblasti. Pod ní následuje kategorie širokých schopností (např. *Fluidní inteligence*, *Rychlost zpracování*), které se promítají do užších oblastí. Nejnižší leží specifické schopnosti, které vysvětlují výsledky v konkrétních testech či úspěch v konkrétních činnostech (např. *Kvantitativní usuzování*, *Kapacita pracovní paměti*).

V rámci této taxonomie náleží test *Lin* faktorům specifických schopností *Vizualizace* (*Vz*) a částečně *Vizuální paměť* (*MV*), které oba spadají pod široký faktor *Vizuální zpracování* (*Gv*).



Obrázek 3. Faktorová struktura C-H-C teorie, (Schneider & McGrew, 2018)

Široká schopnost – Vizuální zpracování (Gv)

Vizuální zpracování (někdy také označované jako *Prostorové schopnosti*) lze definovat jako schopnost používat vnitřní obrazové reprezentace k řešení problémů – vnímání, rozlišování, manipulaci a znovu-vybvavování si podnětů (Schneider & McGrew, 2018). Nejde přitom o operace primárního smyslového vnímání, ale spíše o procesy pozdějšího zpracování

smyslových vjemů. Příkladem může být orientace v mapě nebo schopnost vybavit si v mysli trojrozměrný tvar a určit, jak bude vypadat po otočení nebo rozpůlení.

Podle Carrolla (1993) jde o jeden z nejméně studovaných širokých faktorů C-H-C. Kvůli relativně nižším korelacím s některými studijními a pracovními úspěchy byl ale historicky v praxi větší zájem o *Gf* (*Fluidní inteligenci*) a *Gc* (*Krystalickou inteligenci*). V poslední době se ale zájem o *Gv* zvyšuje a objevují se studie (např. Wai et al., 2009), podle kterých jsou výsledky v *Gv* testech dobrým prediktorem úspěchu v tzv. STEM oborech (věda, technologie, inženýrství, matematika). Současný rozvoj softwaru a virtuálních prostředí si navíc klade čím dál větší nároky na práci s prostorovými koncepty. Existuje proto motivace vyhledávat děti mimořádně nadané v této oblasti a postarat se o to, aby měly k dispozici podporu pro rozvoj svého talentu. Zejména v případech, kdy takové děti současně nevynikají v oblasti verbálních či kvantitativních schopností. Lakin a Wai (2020) popsali, že takoví žáci nemusí nutně dosahovat dobrých školních výsledků a mohou trpět nedostatkem motivace ke studiu.

Úzká schopnost – Vizualizace (Vz)

Podobně jako je *Indukce (I)* centrální úzkou schopností pro *Fluidní inteligenci (Gf)*, je i *Vz* centrální pro *Gv* a je obvykle vnímána jako zástupná proměnná pro celé *Gv*. Jedná se o schopnost vnímat komplexní tvary či vzorce a představit si jejich změněný tvar, např. po rotaci, zúžení, částečném zakrytí (Schneider & McGrew, 2018). Jde také o úzkou schopnost *Gv*, která nejvíce koreluje s obecnou inteligencí. Důvodem je zřejmě komplexnost operace, která je vyšší než u jiných *Gv* podfaktorů, např. *Odhadování délky (LE)* či *Rychlosti rozpoznávání (CF)*.

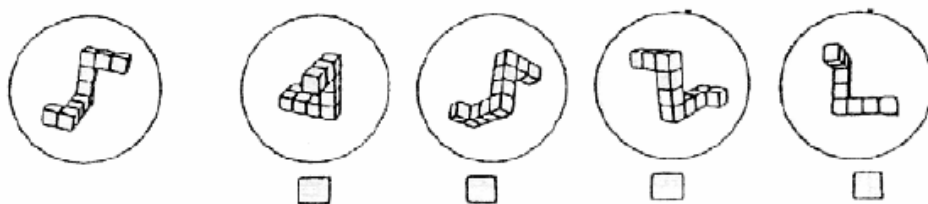
V rámci *Vz* (i obecně *Gv*) lze nalézt facety, které dále kategorizují různé mentální operace. Uttal a kolegové (2013) uvádějí jako nejčastější dělení na statické vs. dynamické prostorové schopnosti a poté operace zaměřené na jeden objekt vs. na orientaci v prostředí. Zatímco většina existujících metod pracuje pouze se statickými podněty, test *Lin* využívá formát počítačového testování k prezentaci animovaných, dynamických podnětů. Právě kvůli nutnosti krátce si zapamatovat pohybuje se objekt, předpokládáme určitý vliv další úzké schopnosti, *Vizuální paměti (MV)*.

Úzká schopnost – Vizuální paměť (MV)

U této schopnosti jde o zapamatování si komplexního obrazce či vzoru po krátkou dobu – méně než 30 sekund (Schneider & McGrew, 2018). Rozdíl od *Kapacity krátkodobé paměti (Gwm)* spočívá v tom, že jde o jeden komplexní obrazec. U krátkodobé paměti jsou to sekvence méně složitých podnětů, kde navíc záleží na pořadí (tamtéž).

Existující testové principy a podobně zaměřené testy

Zájem o měření *Gv* se datuje k sedmdesátým létům. Zřejmě nejznámějším testem prostorových schopností jsou tzv. *Mentální rotace* (Vandenberg & Kuse, 1978). Řešitel zde musí vybrat z různě otočených těles to, které je totožné s obrázkem v zadání, viz obrázek 4 níže. Musí si přitom vytvořit správný mentální obrázek požadovaného tělesa, které si potom v duchu otočí a srovná s nabízenými možnostmi. Čím více stupňů otočení je vyžadováno, tím jsou úlohy těžší. Popis této mentální operace odpovídá definici *Vz*. Většina těchto testů však pracuje s časovým limitem a úlohy jsou natolik rozšířené, že se vyčleňuje specifická úzká schopnost v rámci *Gv* a to *Rychlost rotace (SR)*.



Obrázek 4. Ukázka položky založené na principu testu *Mentálních rotací* (Vandenberg & Kuse, 1978).

Mentální rotace se v různých podobách stále objevují v inteligenčních testech nebo v testech studijních předpokladů, např. IST (Liepmann et al., 2007) či TSP (Masarykova univerzita, 2022). Častěji se ale pracuje se subtesty zaměřenými na obecnější schopnost vizualizace, nejen čistě na rotaci. Ve známém Wechslerově testu pro dospělé (Wechsler, 2008) je *Gv* měřeno v subtestech *Skládání kostek*, *Puzzle*, *Doplňování obrázků* a *Doplňování vah*.

Co se týká testů zaměřených na dětskou populaci, ve Spojených státech nyní existuje metoda přímo zaměřená na vyhledávání nadaných v oblasti *Vizuálního zpracování*. Test *Spatial Test Battery* (Stumpf et al., 2013) je rozdělen na části *Mentální rotace*, *2D projekce těles*, *Perspektiva* a *Vizuální paměť*. Metoda má velmi dobrou prediktivní validitu vůči úspěchu ve STEM oborech, dokonce větší než americké *high stakes* testy SAT (tamtéž). V dalších testech se objevují subtesty *Skládání bloků* a *Puzzle* (WISC-IV; Wechsler, 2014), *Kopírování tvarů*, *Vizuální paměť*, *Skládání tvarů* (DAS-II; Elliott, 2007), *2D rotace*, *3D rotace*, *Skládání tvarů* (WJ-IV; Schrank, 2014) apod. Již z názvů subtestů je zřejmé, že jsou vesměs zaměřené na *Vizualizaci*, často také na *Vizuální paměť*. Protože jde vždy o metody administrované ve formátu tužka-papír, měří statické prostorové schopnosti.

Dle výsledků naší rešerše neexistuje žádná plošně rozšířená standardizovaná metoda, která by využívala animované, dynamické podněty. Naším záměrem bylo využít co nejvíce gamifikačních prvků, které by zvýšily motivaci dětí či snížily případný strach z testování. Proto považujeme inovativní volbu animovaných podnětů za vhodnou.

Časový limit

Většina výše zmíněných testů pracuje s časovým limitem. U testů mentálních rotací je časový limit dokonce součástí definice úzké schopnosti *SR*. V testu *Lin* jsme se rozhodli s časovým omezením nepracovat a to z několika důvodů:

1. Z hlediska obsahové validity je účelem testu *Lin* měřit *Vizualizaci*. V některých taxonomiích C-H-C teorie (např. McGrew, 2009) je přitom *Rychlost usuzování (RE)* samostatnou úzkou schopností. Naším cílem je měřit jednotlivé schopnosti tak, aby byl vliv ostatních schopností na výsledek v testu co nejvíce omezen.
2. Primárním účelem systému Invenio je vyhledávání nadaných žáků. Souhlasíme přitom s názorem některých odborníků (Gridley et al., 2003), že je vhodnější vyhnout se časovým limitům a bonusům za rychlé řešení v testech, u kterých je to možné a rychlost reakce není součástí měřeného konstruktů, a to např. proto, že někteří perfekcionističtí nadaní žáci preferují správnost před rychlostí. Současně se ukazuje, že se kognitivní nadání nemusí vždy projevit v rychlosti zpracování informací (Reams et al., 1990).

- Kromě toho předpokládáme, že časový limit může zvyšovat úzkost z testování. Ukazuje se přitom, že časový limit může v testových situacích znevýhodňovat úzkostné osoby (Onwuegbuzie & Seaman, 2010).

Popis a struktura

V této sekci je představena finální podoba metody *Lin*.

Herní příběh

Hráč se ocitá na planetě, která slouží k zaznamenávání pohybu vesmírných tělech pro účely navigace, a potkává zde astronomku Lin (viz medailonek testu na obrázku 5). Zjišťuje, že počítač se záznamy drah vesmírných těles se porouchal a ukazuje pro každé těleso více než jeden záznam. Rozhodne se proto astronomce Lin pomoci najít a přiřadit správné záznamy.



Obrázek 5. Medailonek testu *Lin a vesmírná navigace*.

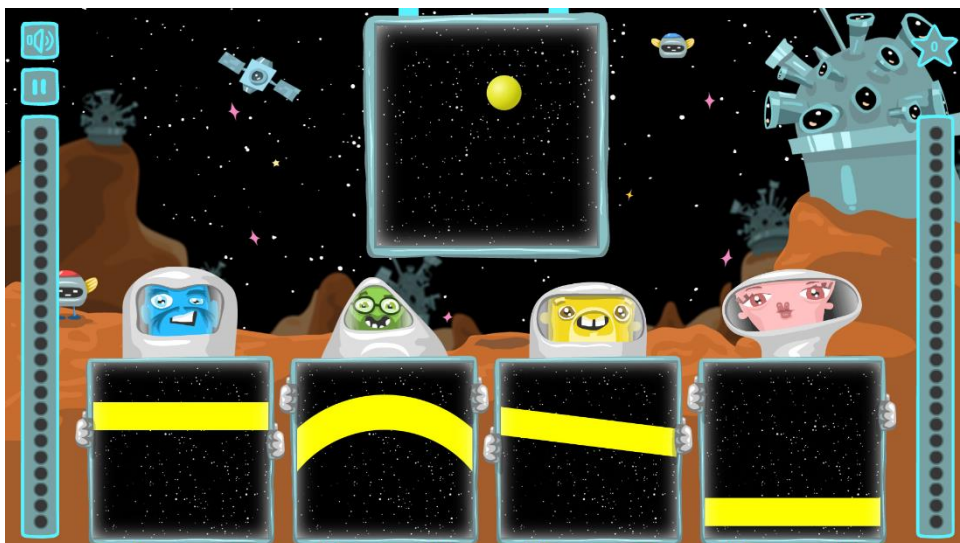
Struktura a jednotlivé prvky testu

Test obsahuje celkem 37 skórovanych položek, rozdělených do tří částí, viz tabulku 2 níže. Po úvodním vysvětlení příběhu následují tyto tři části, vždy v uvedeném pořadí. Nejprve je princip hráči vysvětlen na ukázkové neskórované položce, poté následuje jedna nebo dvě zácvičné neskórované položky. V případě špatné odpovědi dostane hráč zpětnou vazbu s dodatečným vysvětlením principu.

Tabulka 2. Základní přehled jednotlivých částí testu.

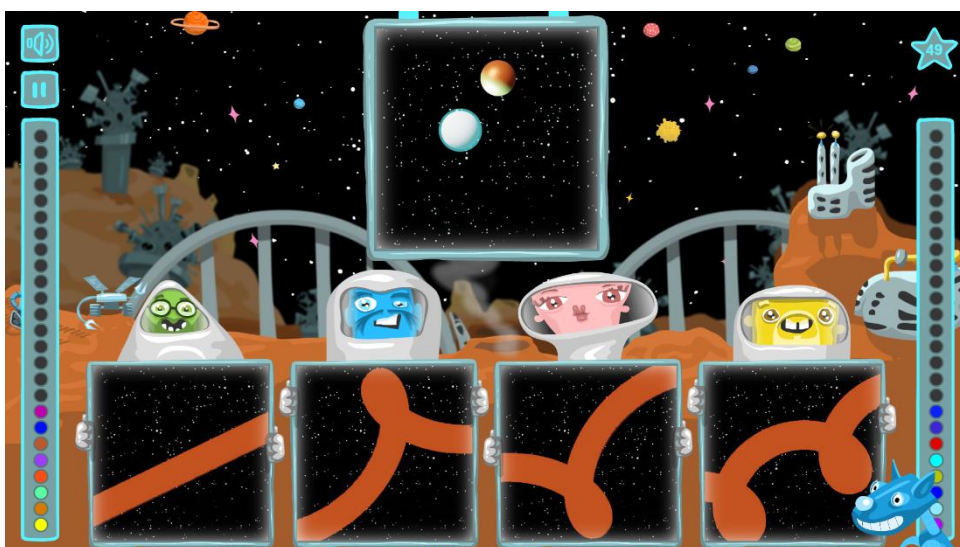
Část	Označení části	Popis principu	Počet skórovanych položek	Počet neskórovanych položek
1.	Jednoduchý pohyb	Je třeba určit dráhu tělesa, které se pohybuje po přímé nebo kruhové dráze	14	3
2.	Složený pohyb	Je třeba určit dráhu satelitu, jehož planeta se sama pohybuje	14	2
3.	Neviditelné těleso	Stejně jako předchozí princip, těleso ovšem samo není vidět. Je třeba orientovat se podle světla, které vyzařuje.	9	2

Centrálním prvkem herního prostředí je velká hlavní obrazovka, na které hráč vidí animace pohybu vesmírných těles (viz obrázek 6 níže). Pod touto obrazovkou vybírá ze čtyř nabízených tvarů dráhy tělesa. V první části testu je na obrazovce vidět vždy jen jeden objekt, dráhy v možnostech mu odpovídají barvou.



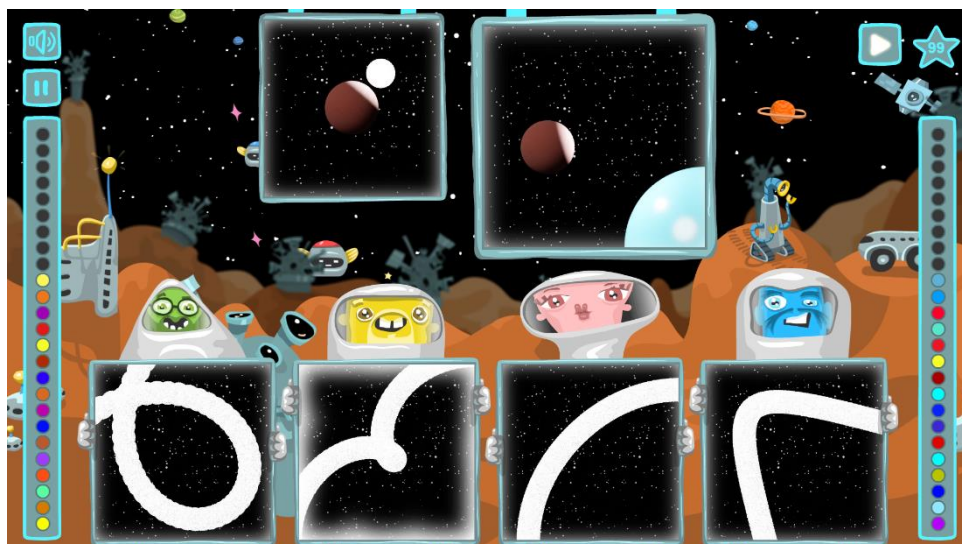
Obrázek 6. Ukázka zácvičné položky první části.

Druhá část testu obsahuje těžší úlohy. Hráč opět určuje dráhu tělesa, nyní se ale jedná o složitější trajektorie, které jsou výsledkem skladby pohybu planety a oběhu jejího satelitu (viz obrázek 7). Těleso, jehož dráhu je potřeba uhodnout, pozná hráč podle stejné barvy. Ve všech úlohách ale jde o objekt, který je satelitem většího objektu.



Obrázek 7. Ukázka zácvičné položky druhé části.

V poslední části testu se objevují úlohy, které jsou obdobou předchozích s tím rozdílem, že hlavní animace na velké obrazovce neukazuje přímo těleso, jehož dráhu je potřeba určit. To je viditelné pouze na přidavné obrazovce vlevo (viz obrázek 8). Těleso vyzařuje silné světlo a hráč určuje dráhu podle něj. Pomáhá mu také informace z menší obrazovky, kde je vidět vzdálenost svítícího satelitu od planety. Pro vyřešení úlohy je tedy nutné zkombinovat informace z obou obrazovek. Naším cílem zde bylo zvýšit obtížnost položek tím, že k vyřešení bude zapotřebí další kognitivní operace – představení si neviditelného objektu na správném místě.



Obrázek 8. Ukázka zácvičné položky třetí části.

Jakmile hráč zaklikne vybranou odpověď, objeví se tlačítko se šipkou pro pokračování. Po jeho stisknutí obdrží zpětnou vazbu ohledně správnosti odpovědi. V případě chyby hra nezobrazuje správnou odpověď.

Kromě příběhu a grafické podoby jsme do testu zahrnuli i další gamifikační prvky. Na bocích obrazovky jsou lišty, ve kterých za každou správnou odpověď přibude symbol planety stejné barvy, jakou mělo těleso v položce. Hráč také sbírá body, které se zobrazují v pravém horním rohu obrazovky. Za každou správnou odpověď získává 3 body, za každou špatnou 1 bod. Tyto body slouží také jako bezprostřední celková zpětná vazba pro děti na konci testu.

Vývoj testu a odpověďové procesy

Na začátku vývoje testu jsme se soustředili hlavně na to, jak nejlépe využít potenciál prezentace podnětů na počítači. Právě prostorové schopnosti nejvíce vybízejí k zakomponování složitějších grafických prvků jako jsou animace a 3D objekty. Po prostudování a zhodnocení množství principů, které se objevují například v testech Woodcock-Johnson, WISC či některých testech studijních předpokladů, jsme zúžili výběr právě na ty principy, které byly potenciálně nejvíce vizuálně vděčné. Jedním z takových principů byl i typ úloh, při kterých je třeba si podle instrukcí představit pohyb znázorněného tělesa. Tento typ úloh nám byl inspirací k vytvoření zcela nového principu. Hráči musí správně zpracovat prezentované animace, vytvořit si představu o dráze objektů a tu potom porovnat s nabízenými možnostmi.

Předpokládáme, že k vyřešení úloh je zapotřebí správně provést několik postupů zapojujících *Vizuální zpracování*. Testovaný musí sledovat pohyb tělesa a vytvořit si vnitřní obraz tohoto pohybu. Musí přitom správně určit referenční rámec pohybu, tj. vůči obrazovce a ne vůči ostatním tělesům v animaci. Musí také správně odhadnout vzdálenosti od krajů obrazovky, případně vzdálenosti od jiných relevantních bodů v poli. Stejně je potřeba odhadnout tvar křivek, poloměr otáček apod. u složitějších drah. U mnoha úloh je také potřeba posoudit, zda je pohyb těles symetrický. Toto vše je nutné uchovat v krátkodobé paměti a integrovat do jednoho výsledného vnitřního obrazu. Ten potom testovaný porovnává s předloženými možnostmi, což vyžaduje použití dalších prostorových schopností, zejména identifikování odlišností, posuzování vzdáleností a tvarů křivek. U těžších úloh předpokládáme, že může být nutné předložené distraktory zpětně „překódovávat“ do animovaného pohybu a posoudit plauzibilitu možnosti.

U třetího typu úloh předpokládáme zapojení stejných procesů jako u předchozích, pouze s tím rozdílem, že je přidán jeden kognitivní krok navíc. Hráč totiž nemůže pracovat přímo

s animovaným objektem, ale musí sis jej nejprve představit či „domyslet“ do animace. Kromě tohoto kroku navíc předpokládáme také zvýšení náročnosti všech následných kognitivních operací, jelikož výstup z prvotní představované animace bude obsahovat více nejistoty ohledně přesné trajektorie tělesa.

Jako jedno z potenciálních rizik pro validitu měření považujeme alternativní strategie řešení. Naším záměrem bylo konstruovat úlohy tak, aby k vyřešení bylo skutečně potřeba následovat výše zmíněný postup. Procesem ověření fungování položek jsme se poté snažili vyselektovat ty položky, které tento záměr ne zcela naplňovaly. Přesto nelze vyloučit, že někteří hráči se svým individuálním stylem řešení budou schopni správnou odpověď odhadnout alternativním postupem, zejména zapojením schopností z oblasti *Fluidní inteligence (Gf)*, tj. odhadnutím správné odpovědi eliminací distraktorů spíše s využitím logické úvahy. Tomuto riziku pro validitu testu plánujeme věnovat pozornost i při budoucích kalibracích a vylepšeních testu.

První pilotní studie

Různými modifikacemi jsme vytvořili 20 položek s širokým spektrem obtížnosti, které jsme zadali 154 dětem napříč ročníky. Sebraná data jsme analyzovali pro ověření použitelnosti principu. Dle faktorové analýzy vykazovala pracovní verze testu jednofaktorovou strukturu a očekávanou zvyšující se úspěšnost ve vyšších ročnících.

Druhá pilotní studie

Na základě nejlépe fungujících položek z předchozí studie jsme vytvořili 70 položek pro druhé pilotní ověření. Tyto jsme rozdělili do dvou skupin po 35 položkách a zadali je dohromady 225 žákům. Z těchto položek jsme opět na základě item-total korelací a indexů infit a outfit vyřadili špatně fungující položky, čímž vznikl set 48 finálních položek pro standardizační studii.

Standardizační studie

Během standardizační studie jsme administrovali 48 položek, včetně 4 zácvičných. Cílem bylo ověřit fungování tohoto finálního setu položek a vytvořit normy pro cílovou populaci dětí od první do šesté třídy ZŠ.

Nejprve uvádíme deskriptivní analýzu položek pomocí Klasické testové teorie, poté analýzy za pomoci Teorie odpovědi na položku. K těmto analýzám jsme použili balíček Mirt (Chalmers, 2012). Pro podrobnější informace k použitému postupu IRT analýzy viz např. de Ayala (2009).

Vzorek

Standardizační studie se zúčastnilo $N = 1689$ žáků z 21 českých základních škol (viz Tabulka 3). Vyřazení z analýz byli ti žáci, kteří nevyplnili ani první dvě celé části testu, nebo jim průchod zabral více času než tři směrodatné odchylky od průměrného času řešení. Popis vzorku podle ročníku se nachází v Tabulce 4.

Tabulka 3. Školy podílející se na standardizaci.

Název školy	Kraj	Velikost sídla*
ZŠ a MŠ Dolní Třebonín	Jihočeský	A
ZŠ Bratčice	Jihomoravský	A
ZŠ Miroslav	Jihomoravský	A
Základní škola a Mateřská škola Čistá	Pardubický	A
ZŠ Malšovice	Ústecký	A
ZŠ Herálec	Vysočina	A
ZŠ a MŠ Krčín	Královéhradecký	B

Základní škola Edvarda Beneše Písek	Jihočeský	C
Základní škola Otto Wichterleho Prostějov	Olomoucký	C
Základní škola Svitavy	Pardubický	C
ZŠ Litomyšl	Pardubický	C
Gymnázium Václava Hlavatého, Louny	Ústecký	C
Základní škola Humpolec	Vysočina	C
ZŠ a MŠ Chalabalova	Jihomoravský	D
ZŠ Edvarda Beneše Opava	Moravskoslezský	D
22. základní škola Plzeň	Plzeňský	D
Základní škola Zlín, tř. Svobody 868	Zlínský	D
2. ScioŠkola, Praha Jarov	Praha	E
Škola Vitae	Praha	E
Základní škola Poznávání	Praha	E
ZŠ Fr. Plamínkové	Praha	E

Pozn. * A – Obce s 1999 a méně obyvateli; B – Malá města 2 000-9 999; C – Střední města s 10 000-49 999; D – 50 000 a více (mimo Prahy); E – Praha.

Tabulka 4. Popisné statistiky standardizačního vzorku.

Ročník	N	Typický věk	% dívek
1	160	7	47
2	157	8	41
3	211	9	49
4	313	10	47
5	358	11	54
6	264	12	50
7	208	13	46
8	18	14	66

CTT položková analýza

V rámci klasické položkové analýzy jsme se zaměřili na obtížnosti položek napříč ročníky, celkovou obtížnost a korelace položek s celkovým skóre. Výsledky analýzy jsou k náhledu v Tabulce 5. Hodnoty představují průměrné výsledné skóre na položku (Hodnota 1 znamená samé správné odpovědi). Korelace v posledním sloupci tabulky jsou biseriální korelace, tj. korelace mezi intervalovou proměnnou celkového skóre a latentním kontinuem za dichotomickými položkami.

Jelikož položky mají vždy čtyři možné odpovědi, nejnížší očekávaná úspěšnost (s ohledem na tipování) by se měla pohybovat okolo 0.25. U korelací bychom očekávali středně silné až silné pozitivní korelace. Na základě této analýzy jsme tedy identifikovali ty položky, které měly buďto nečekaně nízkou úspěšnost (potenciálně špatně fungující distraktory), nečekaně vysokou úspěšnost (triviální položky), nebo nekorelovaly s celkovým skóre. Položky vyznačené tučně jsme z finální verze testu odstranili – na základě této a následujících analýz.

Zároveň jsme ale v testu ponechali i relativně hodně jednoduché položky, které relativně vysoce korelovaly s celkem. Test je určen pro širokou věkovou skupinu a takové položky jsou proto žádoucí, v neposlední řadě pro udržení motivace nejmladších řešitelů.

Tabulka 5. Deskriptivní statistiky jednotlivých položek.

Princip / položky	Obtížnost dle ročníků								Korelace s celkem	
	Celkem	1.	2	3	4	5	6	7		8
Jednoduchý pohyb										
d_p4	0.91	0.76	0.81	0.9	0.93	0.94	0.98	0.98	1	0.61
d_p5	0.83	0.59	0.76	0.75	0.85	0.87	0.91	0.95	1	0.55
d_p7	0.81	0.58	0.69	0.71	0.8	0.88	0.89	0.96	0.83	0.58
d_p8	0.96	0.87	0.92	0.92	0.98	0.98	0.99	0.99	1	0.63
d_p9	0.92	0.79	0.84	0.85	0.91	0.96	0.98	0.98	1	0.6
d_p10	0.97	0.9	0.95	0.94	0.98	0.99	1	1	1	0.74
d_p11	0.74	0.61	0.66	0.67	0.69	0.8	0.8	0.84	0.83	0.44
d_p12	0.89	0.78	0.83	0.82	0.92	0.92	0.94	0.93	0.89	0.46
d_p13	0.91	0.81	0.84	0.86	0.9	0.93	0.97	0.94	1	0.57
d_p14	0.97	0.91	0.96	0.97	0.96	0.99	1	1	1	0.62
d_p16	0.76	0.57	0.65	0.71	0.75	0.81	0.82	0.86	0.94	0.52
d_p18	0.7	0.53	0.53	0.58	0.74	0.76	0.78	0.82	0.89	0.48
d_p19	0.49	0.28	0.32	0.43	0.49	0.53	0.62	0.61	0.83	0.47
d_p20	0.89	0.79	0.76	0.87	0.89	0.91	0.97	0.96	1	0.49
d_p24	0.97	0.85	0.92	0.96	0.98	0.99	1	1	1	0.77
d_p26	0.98	0.92	0.97	0.98	0.99	0.99	0.98	1	1	0.58
Složený pohyb										
d2_p2	0.81	0.61	0.64	0.68	0.84	0.87	0.91	0.96	1	0.66
d2_p4	0.63	0.36	0.44	0.54	0.61	0.68	0.76	0.83	0.89	0.59
d2_p5	0.58	0.42	0.43	0.5	0.52	0.63	0.71	0.74	0.89	0.47
d2_p6	0.79	0.6	0.62	0.71	0.76	0.86	0.93	0.88	0.94	0.62
d2_p7	0.43	0.22	0.32	0.36	0.38	0.43	0.56	0.6	0.72	0.51
d2_p8	0.71	0.53	0.57	0.66	0.69	0.76	0.8	0.81	0.89	0.41
d2_p10	0.81	0.62	0.69	0.73	0.78	0.88	0.9	0.96	0.83	0.61
d2_p11	0.73	0.54	0.64	0.68	0.72	0.79	0.75	0.86	0.89	0.41
d2_p12	0.63	0.52	0.42	0.57	0.59	0.66	0.75	0.79	0.94	0.46
d2_p13	0.46	0.24	0.32	0.31	0.43	0.52	0.6	0.64	0.78	0.53
d2_p14	0.7	0.51	0.5	0.6	0.71	0.76	0.8	0.83	0.89	0.51
d2_p16	0.76	0.54	0.62	0.7	0.78	0.79	0.87	0.88	0.94	0.56
d2_p17	0.77	0.68	0.68	0.7	0.73	0.8	0.84	0.9	0.83	0.5
d2_p18	0.45	0.2	0.28	0.3	0.42	0.54	0.6	0.62	0.83	0.66
d2_p21	0.71	0.52	0.55	0.62	0.68	0.76	0.85	0.88	0.89	0.54
d2_p22	0.39	0.31	0.25	0.32	0.35	0.4	0.45	0.54	0.56	0.42
d2_p24	0.37	0.28	0.23	0.34	0.3	0.4	0.45	0.53	0.72	0.4
d2_p26	0.33	0.16	0.2	0.21	0.27	0.39	0.44	0.5	0.67	0.6
Neviditelné těleso										
d3_p5	0.38	0.19	0.27	0.28	0.34	0.42	0.52	0.53	0.67	0.46
d3_p6	0.46	0.17	0.27	0.31	0.39	0.52	0.62	0.74	0.83	0.59
d3_p7	0.58	0.43	0.45	0.5	0.57	0.63	0.66	0.72	0.72	0.42
d3_p8	0.28	0.16	0.22	0.27	0.25	0.25	0.37	0.39	0.28	0.35
d3_p9	0.22	0.13	0.07	0.11	0.23	0.23	0.28	0.38	0.56	0.39
d3_p10	0.39	0.3	0.25	0.31	0.34	0.44	0.48	0.5	0.56	0.32
d3_p11	0.5	0.53	0.46	0.4	0.48	0.54	0.54	0.54	0.78	0.24
d3_p13	0.37	0.2	0.24	0.27	0.34	0.4	0.48	0.55	0.61	0.44
d3_p14	0.2	0.27	0.22	0.14	0.14	0.2	0.24	0.24	0.22	0.2
d3_p15	0.37	0.21	0.29	0.27	0.33	0.4	0.49	0.52	0.5	0.42
d3_p16	0.48	0.29	0.41	0.4	0.45	0.5	0.57	0.65	0.72	0.41
d3_p17	0.14	0.13	0.13	0.15	0.14	0.14	0.18	0.13	0.11	0.11
d3_p19	0.48	0.39	0.45	0.48	0.47	0.48	0.52	0.53	0.56	0.24
d3_p20	0.66	0.53	0.48	0.62	0.66	0.65	0.76	0.81	0.88	0.43

IRT položková analýza

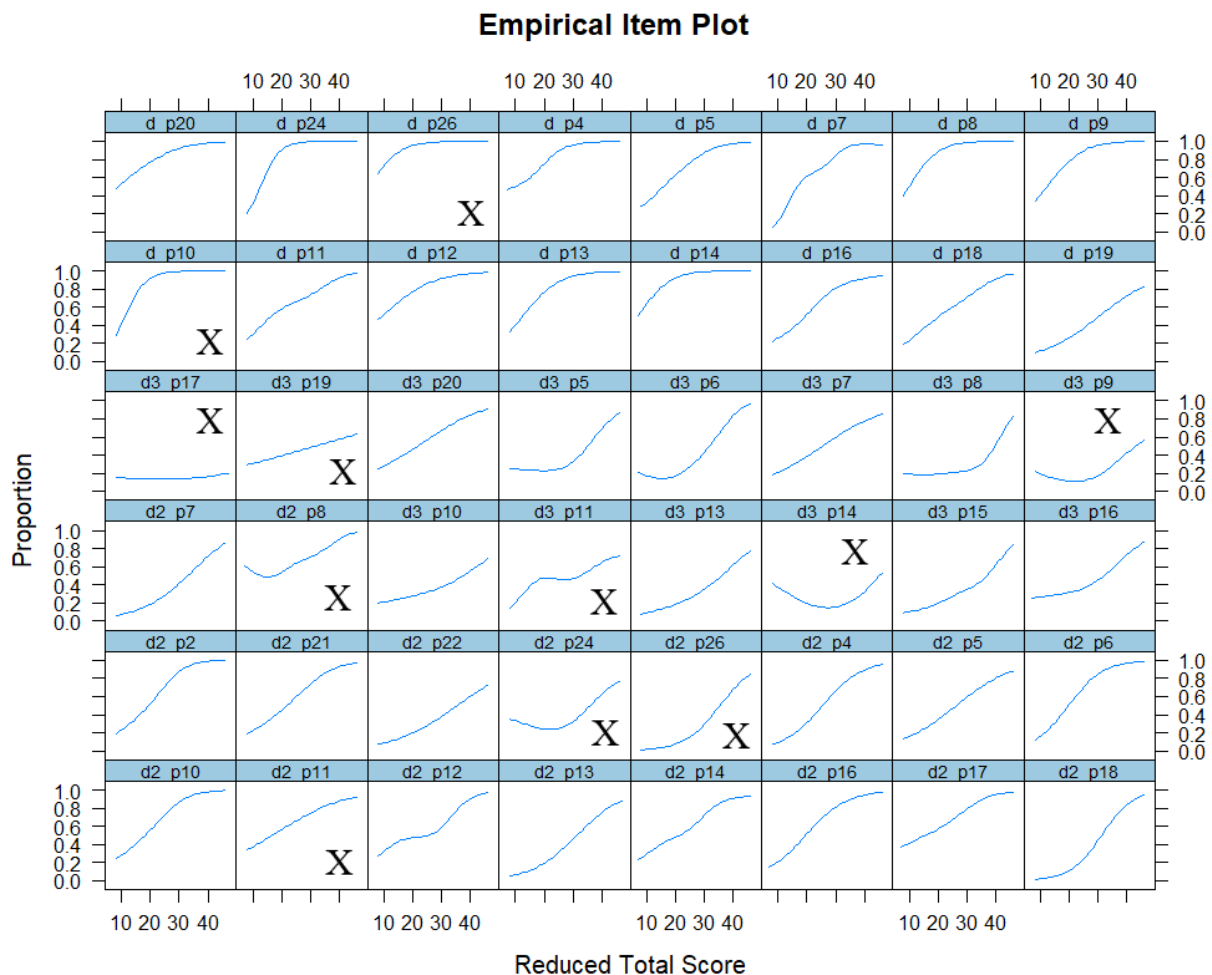
Dále jsme přistoupili k analýze položek pomocí metod z *Item Response Theory* (IRT). Tyto analýzy nabízejí na jedné straně více užitečných informací o fungování položek, na straně druhé je také filozofický základ této teorie měření méně problematický než u CTT. Zejména pokud jde o odůvodnění sčítání skóre jednotlivých položek do celkového skóre. Naším cílem bylo vytvořit metodu, která bude odpovídat Rashovu modelu (Bond et al., 2020), mimo jiného také

kvůli potenciálnímu rozšíření testu o možnost adaptivního testování. Níže proto ověřujeme jednotlivé předpoklady tohoto modelu (např. Sijtsma & Van der Ark, 2018).

Data jsme nejprve analyzovali pomocí Rashova modelu pro všechny participanty společně, se zakompomovaným předpokladem 25% šance na tipnutí odpovědi ze čtyř možností (odhad pomocí *MML*, algoritmem *Expectation Maximization*). Model předpokládá jednodimenzionalitu, lokální nezávislost položek a stejný diskriminační parametr (strmost křivky). Položky se liší pouze obtížností (průsečíky). Tento první model vykazoval adekvátní shodu s daty:

$M_2(1127) = 2295,38$; $p < 0,001$; $RMSEA = 0,025$ $_{90}CI [0,024 ; 0,027]$; $TLI = 0,948$; $SRMSR = 0,050$

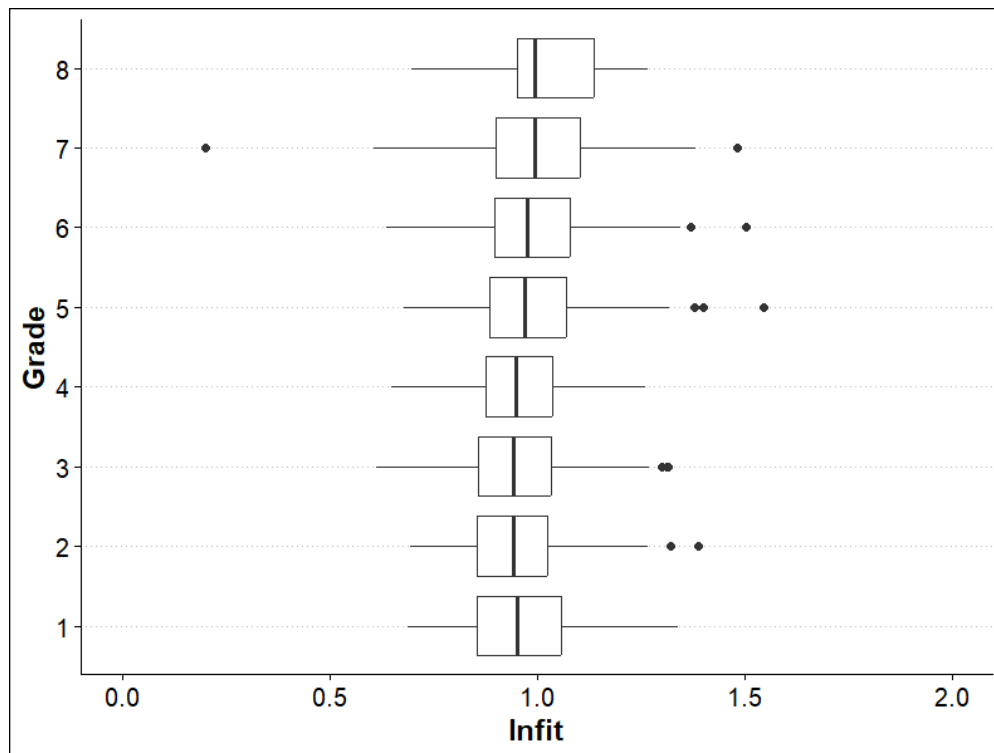
Níže na Obrázku 9 následují vyhlazené pozorované odpověďové křivky položek. Je vidět, že většina položek splňuje předpoklad monotonicity, tj. křivky vztahu schopnosti (celkového skóre) a úspěšnosti (proporce správných řešitelů) jsou po celé délce rostoucí. Toto neplatí pro položky d3_p7, d2_p8, d3_p11, d3_p14 a d2_p24. Závažnost porušení tohoto předpokladu jsme brali v potaz při vyřazování položek. Vyřazené úlohy jsou označeny písmenem „X“.



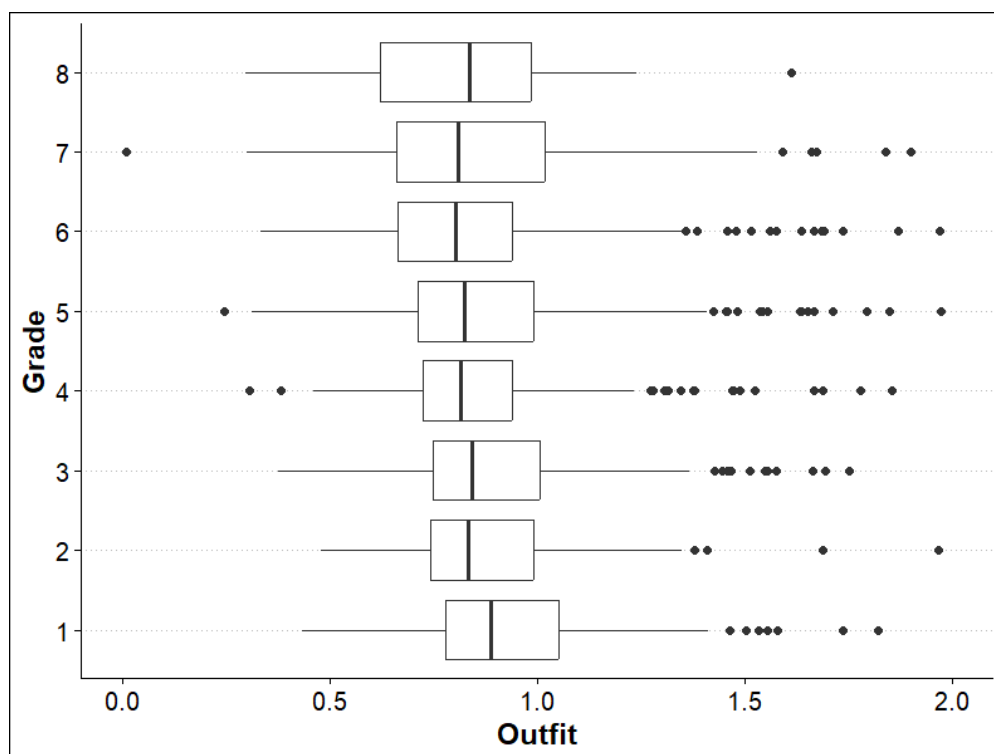
Obrázek 9. Pozorované odpověďové křivky jednotlivých položek.

Před dalšími analýzami položek jsme přistoupili k analýze odpověďových vzorců jednotlivých řešitelů. Některé vzorce odpovědí se mohou z idiosynkrastických důvodů velmi lišit od většiny. Taková data neslouží dobře k účelu ověření fungování položek a v extrémních případech (hodnoty *infit/outfit* větší než 2) mohou výstupy zkreslovat (např. Smith, 1986). Rozložení *infit*

a *outfit* jsou na obrázcích 10 a 11 níže. Mezi ročníky jsme nepozorovali žádné velké rozdíly a v tomto smyslu tedy test funguje podobně pro různé věkové skupiny.



Obrázek 10. Hodnoty *infit* dle ročníku.



Obrázek 11. Hodnoty *outfit* dle ročníku.

Hodnoty *infit* indikují řešitele s nečekanými odpověďmi na položky přibližně odpovídající jejich schopnosti. Tyto hodnoty příliš nevarují, nikdo nepřekročil hranici >2 . Hodnoty *outfit*,

kteře označují zvláštñí odpovědi na velmi lehké či obtížné položky (např. nepozornost, štěstí při hádání), mají větší rozptyl.

Celkově 33 dětí překročilo hranici *outfit* >2 a byly z dalších analýz položek vyřazeny. Tento vzorek se nelišil co do zastoupení jednotlivých škol, věku, ani času řešení. Lišil se celkovým skóre, které bylo o 5 bodů vyšší (35.30 oproti 30.13 u ostatních žáků ve vzorku) a zároveň s nižší variabilitou (minimum správně řešených položek bylo u této skupiny 27 položek oproti 8 u zbytku vzorku). Odpovědi těchto řešitelů zároveň často obsahují chyby u jednoduchých položek v prvním oddílu testu. Vysvětlením by mohla být nepozornost. Hodnoty *infit* ani *outfit* nebyly korelované s úrovní schopnosti.

V následující Tabulce 6 naleznete odhady obtížností položek a jejich shody s Raschovým modelem. Vyřazené položky jsou opět označeny tučně.

Tabulka 6. Položkové parametry a statistiky shody dat s Raschovým modelem.

Položka	<i>b</i>	<i>SE</i>	<i>MSQ outfit</i>	<i>STD outfit</i>	<i>MSQ infit</i>	<i>STD infit</i>	<i>RMSEA</i>	<i>p</i>
Jednoduchý pohyb								
d_p4	-2.58	0.11	0.76	-1.82	0.97	-0.35	0.01	0.33
d_p5	-1.60	0.09	0.83	-2.11	0.96	-0.83	0.02	0.09
d_p7	-1.36	0.09	0.87	-1.82	0.95	-1.25	0.01	0.11
d_p8	-3.57	0.14	0.55	-2.50	0.98	-0.17	0.00	0.79
d_p9	-2.64	0.11	0.77	-1.67	0.98	-0.27	0.00	0.92
d_p10	-3.97	0.16	0.44	-2.84	0.96	-0.25	0.02	0.07
d_p11	-0.74	0.08	0.99	-0.25	1.02	0.59	0.02	0.00
d_p12	-2.22	0.10	1.05	0.44	1.04	0.62	0.02	0.00
d_p13	-2.46	0.10	0.89	-0.83	0.98	-0.30	0.00	0.58
d_p14	-4.05	0.17	0.50	-2.40	1.03	0.24	0.02	0.01
d_p16	-0.98	0.08	1.03	0.44	0.97	-0.80	0.02	0.02
d_p18	-0.47	0.08	0.96	-0.99	0.99	-0.28	0.00	0.55
d_p19	0.97	0.09	0.95	-2.66	0.96	-2.41	0.01	0.28
d_p20	-2.27	0.10	0.99	-0.06	1.03	0.46	0.02	0.01
d_p24	-3.79	0.15	0.44	-3.05	0.95	-0.38	0.02	0.01
d_p26	-4.45	0.20	0.60	-1.61	1.02	0.19	0.01	0.15
Složený pohyb								
d2_p2	-1.44	0.09	0.72	-4.08	0.91	-2.33	0.01	0.28
d2_p4	-0.06	0.08	0.90	-3.03	0.92	-3.94	0.01	0.37
d2_p5	0.36	0.09	0.99	-0.52	0.98	-1.23	0.02	0.05
d2_p6	-1.24	0.08	0.85	-2.27	0.94	-1.81	0.01	0.18
d2_p7	1.36	0.10	0.90	-5.96	0.91	-5.87	0.02	0.08
d2_p8	-0.51	0.08	1.01	0.27	1.03	1.26	0.03	0.00
d2_p10	-1.44	0.09	0.82	-2.49	0.94	-1.45	0.00	0.46
d2_p11	-0.65	0.08	1.05	1.11	1.03	1.19	0.04	0.00
d2_p12	0.01	0.08	0.94	-1.81	0.98	-1.06	0.02	0.06
d2_p13	1.10	0.09	0.93	-4.26	0.93	-5.02	0.02	0.00
d2_p14	-0.51	0.08	0.96	-0.82	0.97	-1.03	0.02	0.01
d2_p16	-1.03	0.08	0.93	-1.14	0.97	-1.08	0.00	0.62
d2_p17	-1.00	0.08	0.95	-0.87	1.00	0.08	0.02	0.09
d2_p18	1.00	0.09	0.83	-9.60	0.85	-10.14	0.04	0.00
d2_p21	-0.58	0.08	0.92	-1.67	0.96	-1.55	0.00	0.99
d2_p22	1.81	0.11	0.95	-2.92	0.96	-2.72	0.02	0.01
d2_p24	1.97	0.12	0.95	-3.00	0.95	-3.08	0.02	0.00
d2_p26	2.01	0.11	0.83	-9.82	0.85	-9.63	0.04	0.00
Neviditelné těleso								
d3_p5	1.79	0.11	0.92	-4.53	0.93	-4.57	0.02	0.01
d3_p6	1.06	0.09	0.87	-7.60	0.89	-7.89	0.02	0.00
d3_p7	0.37	0.09	1.01	0.33	1.01	0.64	0.03	0.00
d3_p8	3.02	0.18	0.90	-4.49	0.91	-4.40	0.02	0.01
d3_p9	3.84	0.28	0.83	-7.16	0.84	-6.84	0.04	0.00
d3_p10	1.97	0.12	1.01	0.47	1.01	0.51	0.01	0.38
d3_p11	1.14	0.10	1.09	4.91	1.08	5.45	0.04	0.00
d3_p13	1.92	0.12	0.93	-3.97	0.94	-3.80	0.00	0.63
d3_p14	5.06	0.63	0.85	-5.75	0.85	-5.64	0.04	0.00

d3_p15	1.97	0.12	0.94	-3.53	0.94	-3.38	0.00	0.66
d3_p16	1.13	0.10	0.99	-0.79	0.99	-0.58	0.01	0.10
d3_p17	13.71	28.18	0.72	-10.91	0.72	-10.91	0.04	0.00
d3_p19	1.33	0.11	1.09	4.90	1.08	5.14	0.04	0.00
d3_p20	-0.15	0.08	1.01	0.42	1.01	0.56	0.03	0.00

Pro ověření lokální nezávislosti položek jsme provedli analýzu residuí pomocí statistiky G^2 (Chen & Thissen, 1997). Žádné hodnoty u ponechaných položek nepřekročily hranici $G^2 = 0.20$. Jednodimenzionalitu testu jsme ověřili skrze analýzu hlavních komponent (PCA) na reziduální matici Raschova modelu. Kromě první komponenty s Eigenvalue 2.09 žádné další nepřekročily hranici 1.50. Loadings této první komponenty byly relativně slabé a distribuované napříč všemi třemi sekcemi testu a nepřikládáme jí proto velkou věcnou významnost. Měřený rys vysvětlil 31 % celkového rozptylu.

Navíc jsme provedli srovnání s alternativním třífaktorovým Raschovým modelem, kde každý faktor odpovídal jedné sekci testu. Tyto faktory byly silně korelované ($r > 0.75$) a model se podle očekávání lépe shodoval s daty ($X^2(5) = 70,30, p < 0,001$), věcně je však tento rozdíl zanedbatelný. Můžeme tedy říci, že výsledný set položek adekvátně odpovídá požadavkům Raschova modelu. Níže uvádíme statistiky shody třífaktorového modelu s daty:

$$M_2(660) = 1247,86; p < 0,001; RMSEA = 0,024 \text{ }_{90}CI [0,022 ; 0,026]; TLI = 0,964; SRMSR = 0,045$$

Provedli jsme také analýzu diferenciálního fungování položek napříč ročníky. Použili jsme SIBTEST (viz Shealy & Stout, 1993), abychom s předstihem získali informaci o potenciálně non-invariantních položkách. Tento postup nevyžaduje odhady latentní schopnosti (a tedy ani specifikaci IRT modelu). Test neodhalil žádné diferenciálně fungující položky.

Níže následuje Tabulka 7 s vyřazenými položkami. Po vyřazení těchto položek a osob s neobvyklým odpověďovým profilem jsme model odhadli znovu. Statistiky shody modelu s daty se mírně zlepšily:

$$M_2(665) = 1307,66; p < 0,001; RMSEA = 0,024 \text{ }_{90}CI [0,023 ; 0,026]; TLI = 0,961; SRMSR = 0,046$$

oproti původním hodnotám:

$$M_2(1127) = 2295,38; p < 0,001; RMSEA = 0,025 \text{ }_{90}CI [0,024 ; 0,027]; TLI = 0,948; SRMSR = 0,050$$

Tabulka 7. Přehled vyřazených položek

položka	infit	outfit	RMSEA	R Celkem
d_p10	0.44	0.44	0.02	0.74
d_p26	0.60	0.60	0.01	0.58
d2_p8	1.01	1.01	0.03	0.41
d2_p11	1.05	1.05	0.04	0.41
d2_p24	0.95	0.95	0.02	0.4
d2_p26	0.83	0.83	0.04	0.6
d3_p9	0.83	0.83	0.04	0.39
d3_p11	1.09	1.09	0.04	0.24
d3_p14	0.85	0.85	0.04	0.2
d3_p17	0.72	0.72	0.04	0.11
d3_p19	1,09	1,09	0,04	0,24

Invariance Raschova modelu

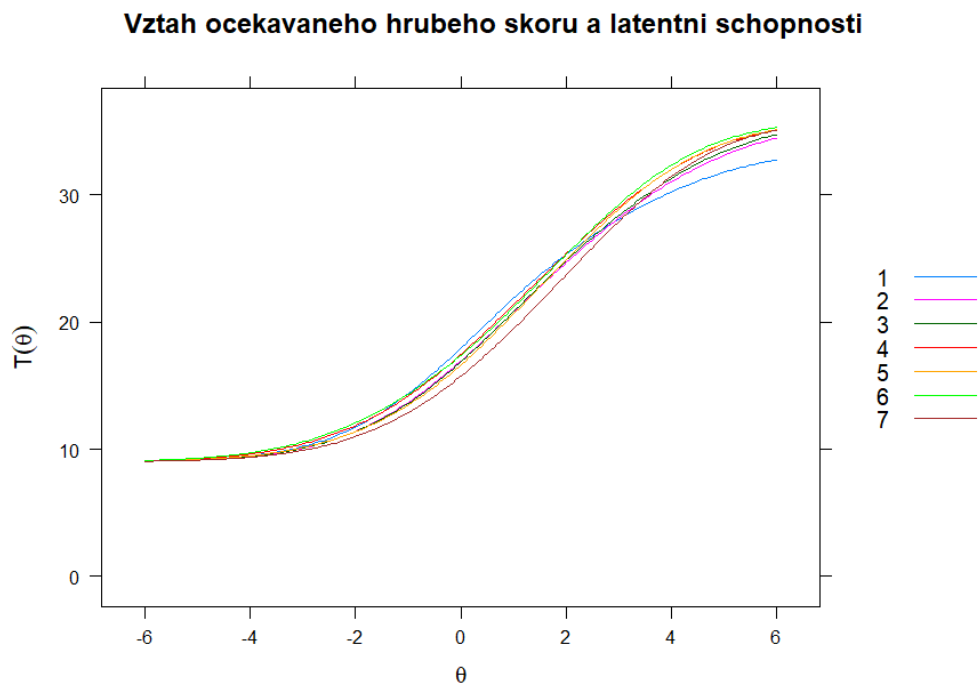
Test je zamýšlen pro použití na celém nižším stupni základních škol. Proto je nutné doložit, že jeho položky fungují akceptovatelně stejně napříč ročníky, tj. jsou invariantní. Pro ověření invariance jsme použili skupinový Raschův model, s volně odhadovanými průměry a rozptyly ročníků a zafixovanými obtížnostmi položek. Z technických důvodů výpočtu jsme ze vzorku odstranili celý osmý ročník (16 dětí) a položku d_p14 s nulovým rozptylem u vyšších ročníků. Shoda tohoto modelu s daty byla adekvátní:

$M_2(4613) = 5433,35$; $p < 0,001$; $RMSEA = 0,011$ 90%CI [0,009 ; 0,012]; $TLI = 0,894$; $SRMSR$ (1.ročník) = 0,086, $SRMSR$ (2. ročník) = 0,091, $SRMSR$ (3. ročník) = 0,081, $SRMSR$ (4. ročník) = 0,071, $SRMSR$ (5. ročník) = 0,060, $SRMSR$ (6. ročník) = 0,080, $SRMSR$ (7. ročník) = 0,096

Dále jsme odhadli skupinový model s volně odhadovanými obtížnostmi pro každý ročník (jako kotevní položky jsme použili d2_p4 a d2_p14). Tento model se s daty shodoval signifikantně lépe, věcně byl však rozdíl zanedbatelný.

$M_2(4409) = 5196,08$; $p < 0,001$; $RMSEA = 0,011$ 90%CI [0,009 ; 0,012]; $TLI = 0,893$; $SRMSR$ (1.ročník) = 0,089, $SRMSR$ (2. ročník) = 0,090, $SRMSR$ (3. ročník) = 0,082, $SRMSR$ (4. ročník) = 0,071, $SRMSR$ (5. ročník) = 0,061, $SRMSR$ (6. ročník) = 0,081, $SRMSR$ (7. ročník) = 0,096

Pro vizualizaci jsou na Obrázku 12 níže k nalezení křivky očekávaného hrubého skóru vůči latentní schopnosti. Je vidět, že ročníky se drží pohromadě, s výjimkou prvňáků. Ti i při vysoké úrovni schopnosti nedosahují takového hrubého skóru jako ostatní. Lze se však domnívat, že k této rozdílnosti existují opodstatněné důvody, např. větší problémy s porozuměním zadání, kapacita pozornosti, problémy provést více myšlenkových operací současně. Celkově invarianci testu považujeme za adekvátní.



Obrázek 12. Charakteristické funkce testu pro jednotlivé ročníky

Post-standardizační (finální) podoba testu

Vzorek

Finální vzorek pro analýzy spojené s post-standardizační verzí testu je uveden v Tabulce 8.

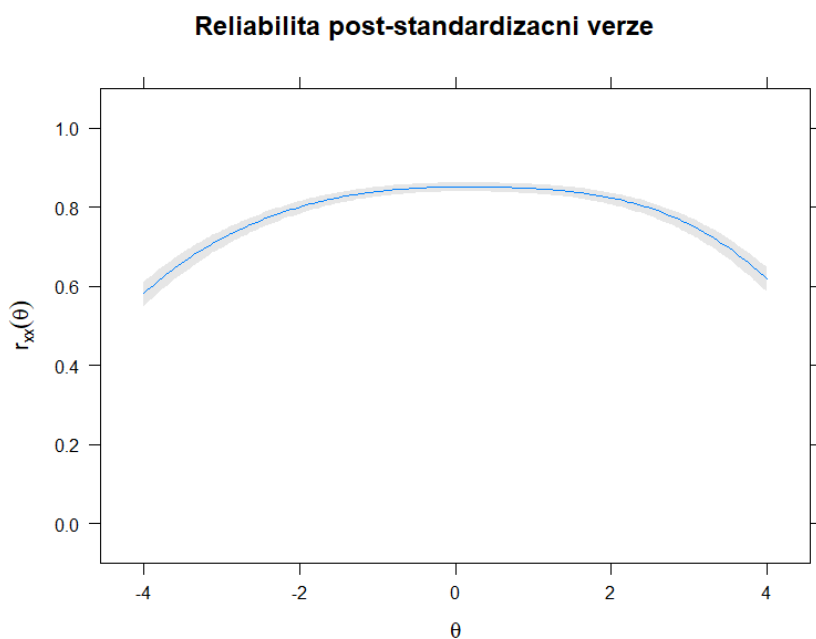
Tabulka 8. Popisné statistiky vzorku pro normy.

Ročník	N	Typický věk	% dívek
1	153	7	48
2	155	8	41
3	201	9	49
4	289	10	47
5	343	11	54
6	264	12	51
7	195	13	46
Celkem	1582		49

Reliabilita

Reliabilitu testu jsme odhadovali pomocí klasické *split-half* reliability a pomocí IRT. *Split-half* reliabilitu jsme vypočítali jako korelaci sudých a lichých položek se Spearman-Brownovou korekcí. Výsledných $r_{xx} = 0.83$ považujeme za přiměřených pro screeningový test.

IRT odhady reliability pro jednotlivé úrovně schopnosti ukazuje Obrázek 13. Odhad marginální reliability pro celý vzorek je $r_{xx} = 0.85$, empirická reliabilita potom $r_{xx} = 0.82$.



Obrázek 13. Vztah reliability testu k úrovni měřené schopnosti.

Normy

Normy spolu s odhady reliabilit pro jednotlivé ročníky uvádíme v tabulkách 9 a 10. Údaje pochází z finálního skupinového IRT modelu, jehož parametry jsou shodné s těmi uvedenými výše v sekci Invariance, pouze bez vyřazení položky *d_p14*. Výsledky jednotlivých testovaných odpovídají standardizovanému rozdílu mezi odhadem úrovně jejich latentní schopnosti a průměrem ročníku. Chyba měření je odhadována specificky pro různé úrovně schopnosti, viz obrázek 13. Normy jsou společné pro chlapce i dívky (k tomuto více v sekci Férovost).

Tabulka 9. Odhady reliability podle ročníků.

Ročník	Odhad reliability		
	Split-half	Marginální IRT	Empirická IRT
1	0.70	0.86	0.71
2	0.66	0.86	0.73
3	0.77	0.86	0.72
4	0.75	0.85	0.75
5	0.75	0.84	0.72
6	0.77	0.81	0.76
7	0.74	0.80	0.70
Celkem	0.83	0.85	0.82

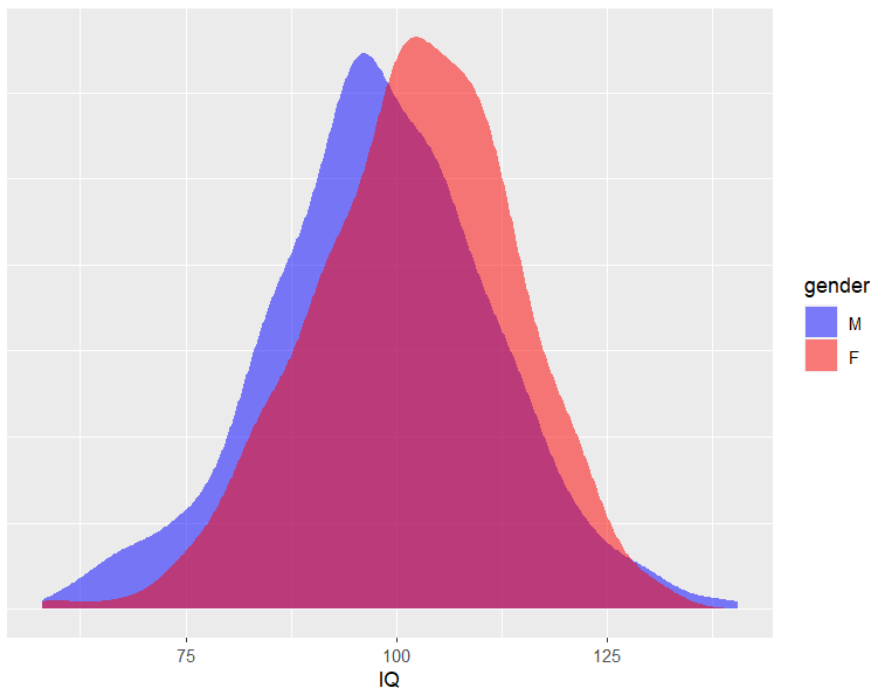
Tabulka 10. Popisné statistiky latentních skóre podle ročníků.

Ročník	Průměrný skóre	SD skóre
1	0.00	1.06
2	0.38	1.02
3	0.76	1.04
4	1.32	1.06
5	1.82	0.95
6	2.39	1.09
7	2.65	0.96

Férovost

Výsledky v testu by neměly být nepřiměřeně ovlivněny charakteristikami, které podle teorie nesouvisí s měřeným konstruktem. Na standardizačních datech jsme ověřovali rozdíly podle pohlaví a podle velikosti sídla.

Na celkovém vzorku jsme pozorovali rozdíl mezi chlapci a dívkami. Průměr dívek byl celkově IQ = 102,04 zatímco u chlapců IQ = 98,06. Rozdíl byl tedy 3,98 IQ bodů ($p < 0,001$), velikost efektu je zde $d = 0,31$. Průměr dívek byl vyšší v každém ročníku, rozdíly ale variovaly: 1. ročník 0,27 IQ, 2. ročník 3,87 IQ, 3. ročník 3,52 IQ, 4. ročník 4,18 IQ, 5. ročník 5,34 IQ, 6. ročník 6,26 IQ, 7. ročník 2,11 IQ. V návaznosti na toto jsme spočítaly rozdíl mezi IRT modelem se stejnými obtížnostmi pro obě pohlaví a volným modelem. Tento rozdíl byl signifikantní, ale věcně nepodstatný ($\chi^2(35) = 105,78, p < 0,001$). Rozložení IQ skóre obou skupin je k vidění na obrázku 14 níže.



Obrázek 14. Rozložení IQ skóre podle pohlaví.

Kromě rozdílu v průměru a modu je z grafu patrný i větší rozptyl skóre chlapců – jsou více zastoupeni jak u nejnižších tak u nejvyšších skóre, což odpovídá obecnému rozložení IQ v populaci. V oblasti mimořádného nadání (dvě směrodatné odchylky od IQ 100) se ve standardizačním vzorku nachází 36 dětí, z toho 15 dívek a 21 chlapců. Celkově považujeme férovost s ohledem na pohlaví za akceptovatelnou. Budeme se však tomuto bodu dále věnovat při budoucích kalibracích testu.

Nalezli jsme také rozdíly mezi průměrnými skóre dětí ze sídel různé velikosti. Děti z Prahy skórovaly zhruba o 5 bodů výše než děti z menších obcí, což odpovídá rozdílu pozorovanému i u dalších Invenio testů. Podobné rozdíly mezi regiony jsou známý, komplexní fenomén, jemuž je v ČR věnována pozornost výzkumníků (např. Prokop, 2019; Šafr et al., 2022). Férovost tohoto testu vůči dětem z různě velkých sídel považujeme za akceptovatelnou.

Validita

Validitu metody plánujeme ověřovat prostřednictvím souběžných a diferenciálních validit při budoucím zadávání testu společně s ostatními diagnostickými hrami baterie Invenio. Dosud proběhla jedna studie s omezeným vzorkem ($N=86$) a pilotážní verzí testu, tj. bez vyřazených položek a za použití hrubých skóre. Tato studie (Prílepková, 2022) zmiňuje korelaci hrubých skóre Lin s oddílem prostorových schopností metody Woodcock-Johnson IV ($r = 0,44$; $p < 0,001$). Pozdější re-analýzou těchto dat s použitím čistých IQ skóre jsme došli k identickému výsledku.

Zároveň jsme také pro účely ilustrace divergentní validity spočítali korelaci se subtestem kvantitativního uvažování ($r = 0,35$; $p < 0,001$) a se subtestem krátkodobé vizuální paměti ($r = 0,31$; $p < 0,001$) baterie WJ IV. Pokud se tento vztah potvrdí v rámci budoucích validizačních studií, naše domněnky o možném vlivu krátkodobé vizuální paměti na úspěch v tomto testu budou vyvráceny.

Poděkování

Rádi bychom poděkovali Rebece Přílepkové za ověření vztahu metody Lin s prostorovými subtesty baterie Woodcock-Johnson IV. Vřelé díky patří také Technologické agentuře ČR, která vývoj testu finančně podpořila v rámci Programu na podporu aplikovaného společenskovědního a humanitního výzkumu, experimentálního vývoje a inovací ÉTA v projektu s názvem Komplexní skriningový diagnostický systém GIS s číslem TL01000494. Bez této podpory by metoda nemohla vzniknout.

Literatura

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Psychological Association.
- Carroll. J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511571312>
- Chalmers, R. P. (2012). Mirt: A multidimensional item response theory package for theREnvironment. *Journal of Statistical Software*, 48(6). <https://doi.org/10.18637/jss.v048.i06>
- Elliott C. D. (2007). *Differential Ability Scales—Second Edition (DAS-II)* [Assessment Instrument]. The Psychological Corporation.
- Gridley. B. E., Norman. K. A., Rizza. M. G., & Decker. S. L. (2003). Assessment of gifted children with the Woodcock Johnson III. In F. A. Schrank & D. P. Flanagan (Eds.). *WJ III clinical use and interpretation: Scientist-practitioner perspectives* (pp. 285–317). Academic Press. <https://doi.org/10.1016/B978-012628982-4/50010-6>
- Hawkins, G. E., Rae, B., Nesbitt, K. V., & Brown, S. D. (2013). Gamelike features might not improve data. *Behavior research methods*, 45(2), 301-318
- Horn. J. L., & Noll. J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan. J. L. Genshaft. & P. L. Harrison (Eds.). *Contemporary Intellectual Assessment: Theories. Tests. and Issues* (pp. 53–91). Guilford Press.
- Kiili, K., & Ketamo, H. (2017). Evaluating cognitive and affective outcomes of a digital game-based math test. *IEEE transactions on learning technologies*, 11(2), 255-263. <https://doi.org/10.1109/TLT.2017.2687458>
- Liepmann. D., Beauducel. A., Brocke. B. & Amthauer. R. (2007). I-S-T 2000 R. Intelligenz-Struktur-Test 2000 R (2.. erweiterte und überarbeitete Aufl.). Göttingen: Hogrefe.
- Masarykova univerzita (2022). *Minulé verze Testu studijních předpokladů | Masarykova univerzita*. [online] Masarykova univerzita. Available at: <<https://www.muni.cz/uchazeci/bakalarske-a-magisterske-studium/minule-verze-tsp>> [Accessed 19 June 2022].
- McGrew. K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*. 37. 1 – 10. <https://doi.org/10.1016/j.intell.2008.08.004>
- Onwuegbuzie. A. J., & Seaman. M. A. (1995). The effect of time constraints and statistics test anxiety on test performance in a statistics course. *The Journal of Experimental Education*. 63(2). 115-124. <https://doi.org/10.1080/00220973.1995.9943816>
- Prokop, D. (2019). *Slepé skvrny: o chudobě, vzdělávání, populismu a dalších výzvách české společnosti*. Host.
- Reams. R., Chamrad. D., & Robinson. N. M. (1990). The race is not necessarily to the swift: Validity of WISC-R bonus points for speed. *Gifted Child Quarterly*. 34(3). 108-110. <https://doi.org/10.1177/001698629003400304>

- Šafr, J., Gorčíková M., & Simonová, N. (2022). „Nerovnosti při přechodu na osmileté gymnázium: panelová studie žáků pražských a středočeských škol.“ in R. Džambazovič, D. Gerbery, J. Sopóci (eds.). *Vzdelanostné nerovnosti v optike sociológie. Teoretické východiská a empirické poznatky*. Bratislava: STIMUL.
- Schneider. W. J., & McGrew. K. S. (2018). The Cattell-Horn-Carroll Theory of Cognitive Abilities. In D. P. Flanagan. & E. M. McDonough (Eds.). *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th edition) (pp. 73-163). The Guilford Press.
- Schrank. F. A., McGrew. K. S., & Mather. N. (2014). Woodcock–Johnson IV Tests of Cognitive Abilities [Assessment Instrument]. Riverside.
- Stumpf. H., Mills. C. J., Brody. L. E., & Baxley. P. G. (2013). Expanding Talent Search Procedures by Including Measures of Spatial Ability: CTY's Spatial Test Battery. *Roeper Review*. 35(4). 254-264. doi:10.1080/02783193.2013.829548
- Uttal. D. H., Meadow. N. G., Tipton. E., Hand. L. L., Alden. A. R., Warren. C., et al. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*. 139(2). 352–402.
- Vandenberg. S., & Kuse. A. (1978). Mental Rotation. a Group Test of Three-Dimensional Spatial Visualization. *Perceptual and Motor Skills*. 47. 599-604.
<http://dx.doi.org/10.2466/pms.1978.47.2.599>
- Wai. J., Lubinski. D., & Benbow. C. P. (2009). Spatial ability for STEM domains: Aligning over fifty years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*. 101. 817–835.
- Wechsler. D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition* [Assessment Instrument]. Pearson.
- Wechsler. D. (2014). *Wechsler Intelligence Scale for Children (5th ed.)* [Assessment Instrument]. Pearson.