

Test Deduktivního usuzování Elliot a nečekání návštěvníci – technický manuál – verze 1.0 (9.2.2022)

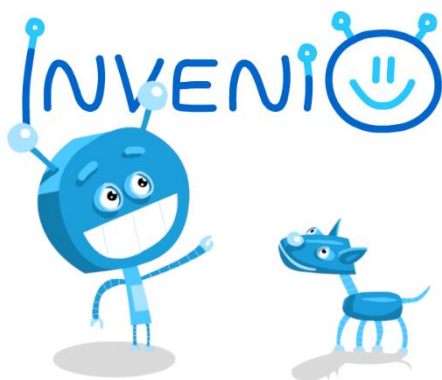
Aktuální verzi citujte jako: Ťápal A., Jabůrek, M., Palíšek, P., & Portešová, Š. (2022). *Test deduktivního usuzování Elliot a nečekání návštěvníci - technický manuál, verze 1.0*. Invenio. Dostupné z www.invenio.muni.cz.

Psychodiagnostický systém Invenio

Úvodní představení

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 a hravosti a redukci případných obav u testovaných dětí.

Níže popsaný test s názvem Triton a hladový oceán (zkráceně *Triton*) je jednou z řady diagnostických her, které jsou součástí systému Invenio. V aktuální podobě je určen žákům 3.–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.

Účel a možnosti použití

Systém Invenio je screeningovým nástrojem, který je zaměřen na jednotlivé úzké schopnosti v rámci inteligenční teorie C-H-C. Může proto být 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 je proto 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 jeden 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 verbálních kategorií, které jsou odvozené od bodové hodnoty standardního skóre (s průměrem 100 a směrodatnou odchylkou 15). Tyto verbální kategorie slouží uživatelům pro snadnější interpretaci výsledků. Jednotlivé verbální kategorie s příslušným rozmezím standardních skóre jsou k dispozici v následující Tabulce 1.

Tabulka 1. Verbální kategorie a jejich vztah ke standardnímu skóru.

Verbální 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.

Elliot a nečekání návštěvníci



Obrázek 3. Medailonek testu Elliot a nečekání návštěvníci.

Východiska a zaměření

Teoretická východiska

Jak již bylo uvedeno výše, systém Invenio a jeho jednotlivé testy vychází primárně z teorie inteligence C-H-C.

C-H-C teorie inteligence

C-H-C (Cattell–Horn–Carroll) teorie (např. Schneider & McGrew, 2018) je integrací dvou předních psychometrických přístupů k inteligenci – Cattell-Hornovy Gf-Gc teorie (např. Horn & Noll, 1997) a Carrollovy teorie tří vrstev (v angl. The Three-Stratum Theory of Cognitive Abilities; Carroll, 1993). Představuje ucelenou taxonomii kognitivních schopností, které rozděluje do tří hierarchických vrstev na základě jejich obecnosti. První vrstvu tvoří velké množství vysoce diferencovaných schopností (tzv. úzké schopnosti neboli narrow abilities) jako je např. *Slovní plynulost*, *Asociační paměť*, *Rychlost usuzování* apod. Tyto dílčí schopnosti jsou ve druhé vrstvě sdružovány do schopností obecnějších, jako je *Krátkodobá pracovní paměť* či *Fluidní inteligence* (tzv. široké schopnosti neboli broad abilities). Třetí, nejvyšší vrstvu, tvoří jediná obecná schopnost obvykle označována jako „g“ faktor, případně jako *Obecná inteligence*.

Označení jednotlivých schopností v rámci teorie C-H-C se může napříč odbornými zdroji a verzemi teorie lišit (např. schopnost Gc bývá označována jako *Krystalická inteligence*, *Porozumění-znalosti* apod.). Současně jsou názvy schopností označovány zkratkami, které jsou naopak konzistentní a mohou proto sloužit jako jednoznačný identifikátor. Z těchto důvodů uvádíme v dalším textu vždy také zkratky úzkých i širokých schopností.

V rámci C-H-C teorie je test *Elliot* zaměřen na širokou schopnost s názvem *Fluidní inteligence* (Fluid reasoning; Gf) a v rámci ní na úzkou schopnost s názvem *Deduktivní usuzování* (Deductive reasoning nebo také General sequential reasoning; RG).

Široká schopnost – Fluidní inteligence (Gf)

Fluidní inteligence představuje jádro toho, co bylo a stále je tradičně pokládáno za inteligenci. Ačkoliv je Gf v dnešní době v rámci teorie C-H-C považována pouze za jednu z řady širokých schopností, je možné ji označit jako základní kámen této teorie. Výsadní postavení této schopnosti se projevuje také na úrovni praktické diagnostiky. Podle Hunta (2011) patří k nejčastěji hodnoceným schopnostem v testech inteligence.

Fluidní inteligence označuje způsob myšlení, který využíváme při setkání s novými úkoly, jejichž řešení nemáme zautomatizované a u kterých není možné se spolehnout na dříve naučené postupy či předchozí znalosti. Tento typ myšlení zahrnuje celou řadu rozličných procesů jako je chápání podstaty konceptů a podobností mezi nimi (např. *co mají společného pes, kočka a kráva*), rozpoznávání vztahů mezi podněty (např. při analogiích typu – *matka k babičce se má stejně jako otec k dědečkovi*), vyvozování závěrů, reorganizaci, transformaci a propojování informací (např. vyvození pravidla, podle kterého je uspořádána číselná či obrázková řada a doplnění chybějícího prvku), tvorbu hypotéz a jejich potvrzování či vyvracení apod.

Obecně lze tuto schopnost chápat jako logické usuzování zaměřené na řešení nových problémů, ve kterém se primárně uplatňují dva procesy - indukce a dedukce.

Úzká schopnost – Deduktivní usuzování (RG)

Deduktivní usuzování představuje schopnost vyvodit závěry z předem daných pravidel (tzv. premis). Podstatou je schopnost vyjít z těchto pravidel a v rámci jednoho či několika a navazujících kroků vytvořit závěr, který lze využít k řešení problému. S ohledem na to, že se jedinec musí při řešení řídit danými podmínkami, bývá deduktivní usuzování někdy označováno jako „*Schopnost aplikovat pravidlo*“ (Rule application). Dalším termínem, který se používá k označení dedukce v rámci C-H-C teorie, je „*Obecné sekvenční usuzování*“ (General sequential reasoning), které zdůrazňuje posloupnost jednotlivých kroků (vyvozování dílčích závěrů) při řešení složitějšího úkolu.

Dedukce souvisí s induktivním usuzováním, které bývá obvykle vymezeno jako schopnost odhalit na základě pozorování dílčích prvků a jejich projevů skryté obecné pravidlo či princip, kterým se chování těchto prvků řídí (např. Schneider & McGrew, 2018). S ohledem na tuto definici se indukce nazývá také jako vyvozování pravidel. Oproti deduktivnímu usuzování se tedy jedná o uvažování v opačném směru. Další rozdíl v obou procesech spočívá v tom, že při procesu deduktivního usuzování nejsou objevovány žádné nové informace, protože všechny už jsou (alespoň implicitně) uvedené v rámci premis (Evans, 2005).

Na rozdíl od jiných testů fluidní inteligence (viz dále) je při hře Elliot a nečekání návštěvníci výhradně uplatňováno deduktivní usuzování, tedy schopnost vyjít z prezentovaných podmínek (premis) a vyvodit z nich logický závěr. Tento obecný princip dále zahrnuje dílčí mentální operace, které jsou nezbytné pro správné řešení úloh. Konkrétním příkladem je schopnost kategorizace, tzn. správné zařazení dílčího prvku do obecnější kategorie (např. pochopení, že premisa týkající se všech zelených ufonů se týká i konkrétního zeleného uфона s jedním okem), případně průnik těchto kategorií (respektive množin, viz následující kapitola).

Existující testové principy a podobně zaměřené testy

Dedukce v rámci inteligenčních baterií

Většina komplexních inteligenčních baterií neobsahuje testy přímo zaměřené na *Deduktivní usuzování*. V manuálu běžně používaných metod jako je např. *Kaufman Assessment Battery for Children II (KABC-II; Kaufman & Kaufman, 2004)* či *Reynolds Intellectual Assessment Scales II (RIAS-2; Reynolds & Kamphaus, 2015)* se autoři při popisu testů, při kterých je uplatňováno logické myšlení, soustředí spíše na obecnější úroveň měření schopnosti a odkazují tedy na *Fluidní inteligenci*. *Deduktivní usuzování* pak obvykle zmiňují jen v rámci výčtu jako jeden ze dvou hlavních procesů fluidní inteligence společně s *Induktivním usuzováním*. Jedná se přitom obvykle o subtesty zaměřené na doplňování logické řady, výběr obrazce, který nepatří mezi ostatní apod. Z povahy testových principů těchto testů je zřejmé, že procesy myšlení, které jsou uplatňovány při řešení, jsou primárně induktivní povahy. Někteří autoři však poukazují na to, že v těchto typech úkolů se kromě indukce uplatňuje i dedukce a oba typy myšlení od sebe může být obtížné odlišovat. Například Schrank (2016) analyzuje problematiku distinkce mezi deduktivním a induktivním usuzováním v rámci testu *Number Series* v baterii *Woodcock-Johnson IV - Tests of Cognitive Abilities (WJ-IV; McGrew et al., 2014)*. Tento test je založen na principu doplňování logických číselných řad, což je proces, který vyžaduje primárně induktivní úsudek. V procesu usuzování lze však podle Schranka spatřovat také prvky dedukce. Testovaný musí nejprve pomocí indukce vyvodit pravidlo dané číselné řady a poté na základě již vyvozeného pravidla deduktivně určit, jaký prvek bude v řadě následovat. Totéž uvádí Carroll (1993), podle kterého každý na indukci zaměřený úkol obsahuje na závěr alespoň jeden krok v podobě deduktivního usuzování.

Dle manuálů nejčastěji používaných inteligenčních testových baterií¹ jsou přímo na *Deduktivní usuzování* zaměřeny pouze dva testy. Prvním z nich je test *Figure Weights* z baterie *Wechsler Intelligence Scale for Children 5 (WISC-V; Wechsler, 2014)*, který je však dle autorů kromě dedukce zaměřen také na *Kvantitativní usuzování*. Druhým je test *Analysis-Synthesis*, který se objevuje v bateriích *Woodcock-Johnson* (např. v poslední edici *WJ-IV*). Úkolem probanda je na základě podmínek (klíče) rozhodnout, jaká barva vznikne kombinací dvou či více jiných barev. V obtížnějších úlohách je nutné postupovat sekvenčně ve více krocích za sebou a kombinovat informace z více různých podmínek. Proces uplatňovaný při řešení těchto úloh se dle našeho názoru velmi blíží tomu, jak je v rámci teorie C-H-C vymezeno *Deduktivní* (neboli *Obecné sekvenční*) *usuzování*.

Absence testů přímo zaměřených na deduktivní usuzování v rámci aktuálně používaných inteligenčních testových baterií je zřejmá. Rozhodli jsme se proto vytvořit nový testový princip, který by stejně dobře

¹ Dle Bensona a kol. (2019) se v praxi školních psychologů v USA za rok 2017 jedná o *WISC-V* (Wechsler, 2014), *DAS-2* (Elliot, 2007), *WJ IV* (McGrew et al., 2014), *KABC-2* (Kaufman & Kaufman, 2004) a *RIAS-2* (Reynolds & Kamphaus, 2015).

zachycoval sekvenční povahu *Deduktivního usuzování* jako v případě testu *Analysis-Synthesis* baterie *WJ-IV* a zároveň by byl lépe uplatnitelný v rámci elektronické formy administrace s herními prvky.

Deduktivní úkoly v rámci výzkumných studií

Kromě testových baterií popsaných výše je možné nalézt úkoly na *Deduktivní usuzování* jako součást výzkumných studií (např. Caparos & Blanchette, 2017; Politzer & Bourmaud, 2002; Rodriguez-Moreno & Hirsch, 2009). Johnson-Laird (1999) v tomto smyslu rozlišuje čtyři typy úkolů. Za prvé jsou to úkoly využívající spojky a, nebo, ale ne, jen a pouze když. Dále úkoly založené na podmiňování (pokud – pak – a proto)². Třetí typ úkolů je zaměřený na usuzování na vztahy³. Čtvrtý typ využívá sylogismy a kvantifikátory⁴. Tyto úkoly (kromě usuzování na vztahy) se staly inspirací pro jednotlivé testové principy metody *Elliot*, které jsou blíže popsány v Tabulce 2.

Výroková logika

Deduktivní usuzování je také úzce propojeno s výrokovou logikou, která je v podstatě formou kategorických sylogismů. Tzn. v jejím rámci jsou (deduktivně) odvozovány úsudky na základě premis (např. Raclavský, 2015). Pro účely testu *Elliot* jsou zásadní tzv. výrokové spojky, které určují vztahy mezi dvěma výroky (viz první typ úkolů Johnson-Lairda). Jedná se především o konjunkci (výroky jsou spojeny spojkou „a“ – závěr je pravdivý pouze pokud jsou oba výroky pravdivé), exkluzivní (neboli ostrou) disjunkci (výroky jsou spojeny spojkou „bud', anebo“ – závěr je pravdivý, když je pravdivý právě jeden z výroků) a ekvivalenci (výroky jsou spojeny spojkou „právě tehdy když“ – závěr je pravdivý, pokud jsou oba výroky pravdivé nebo oba nepravdivé).

Relevantní je v tomto kontextu a pro test *Elliot* také teorie množin. Množina představuje soubor několika prvků, které spojuje určité pravidlo. Pro účely testu jsou důležité vztahy mezi množinami, např. podmnožina představuje množinu, jež je součástí jiné, větší, nadřazené množiny (Cunningham, 2016). Důležité jsou také množinové operace jako např. průnik, kdy mají množiny určitou část společnou.

Einsteinovy hádanky

Základní principy společné s úkoly Johnson-Lairda a výrokové logiky se také objevují v úkolech známých jako Einsteinovy hádanky (Einstein's Riddle), někdy označované jako Zebry (Zebra puzzle). Ty se obvykle využívají ve formě logických hádanek pro děti i dospělé (např. Brož, 2007; Stangroom, 2009), ale v ojedinělých případech se objevují také jako součást výkonových testů (v ČR např. v rámci testů *Obecných studijních předpokladů* společnosti SCIO, kde tyto úlohy obvykle spadají do části testů označované jako analytické myšlení). Ve své původní podobě⁵ je podstatou úkolu situace s několika prvky, které jsou různým způsobem propojeny – např. *existují tři kamarádi, z nichž každý má právě jedno domácí zvíře a jeden dopravní prostředek*. Součástí zadání jsou také výchozí tvrzení, o kterých víme, že jsou pravdivá (premisy) – např. Petr má morče, Karel má auto a Ondra nemá psa. Cílem je vyjít z těchto základních premis a v postupných krocích (sekvencích) odvodit zbývající vztahy mezi prvky.

Integrace východisek do testu deduktivního myšlení

Všechny výše uvedené principy se staly inspirací při tvorbě testu *Elliot*. Hlavním východiskem byl obecný princip výrokové logiky a *Deduktivního usuzování*, tzn. sekvenční vyvozování závěrů na základě premis. Kromě tohoto základního principu pak byly využity dílčí prvky jako např. část výrokových spojek z výrokové logiky, princip přiřazování prvků v případě Einsteinových hádanek apod. V zásadě však test jako celek (včetně většiny dílčích principů) vznikal nově a jeho fungování bylo třeba ověřit v několika navazujících studiích (viz sekce *Pilotní verze č. 1-3*). Jednou z velkých změn oproti

² Např. Pokud Petr chytí rybu, bude ji mít k večeři. Petr chytil rybu. Bude mít rybu k večeři?

³ Např. Eva je lepší než Katka. Anička je horší než Katka. Kdo je nejhorší?

⁴ Např. Někteří politici jsou Angličani. Všichni Angličani pijí čaj. Proto někteří politici pijí čaj.

⁵ Postupně dochází k tomu, že se význam termínu Einsteinovy hádanky či Zebry rozšiřuje a bývají tak označovány i jiné typy úkolů, v rámci kterých je nutné deduktivně vyvozovat závěry z premis.

uvedeným inspiračním zdrojům je například fakt, že veškeré premisy jsou s ohledem na věk probandů ve hře prezentovány neverbálně ve formě grafických piktogramů.

Popis a struktura

V následující Tabulce 1 je zachycen proces vývoje testu a navazujících studií. V této sekci je představena finální podoba metody, která byla použita při standardizaci. Postupný vývoj testu je popsán níže.

Tabulka 1. Časová osa vývoje testu a navazujících studií.

Vývojová fáze	Časové období
Pilotní verze 1 – papírová verze s provizorní grafikou	Listopad, 2016 – Červen, 2017
Pilotní verze 2 – počítačová verze s provizorní grafikou	Červen, 2017 – Červenec, 2017
Pilotní verze 3 a studie Pálkové (2018) – počítačová verze s finální grafikou	Červenec, 2017 – Únor, 2018
Standardizační verze 4 – vývoj	Únor, 2018 – Říjen, 2018
Standardizační verze 4 – sběr dat	Říjen, 2018 – Prosinec, 2018
Studie Wranové (2020) – verze 5	Říjen, 2019 – Duben, 2020

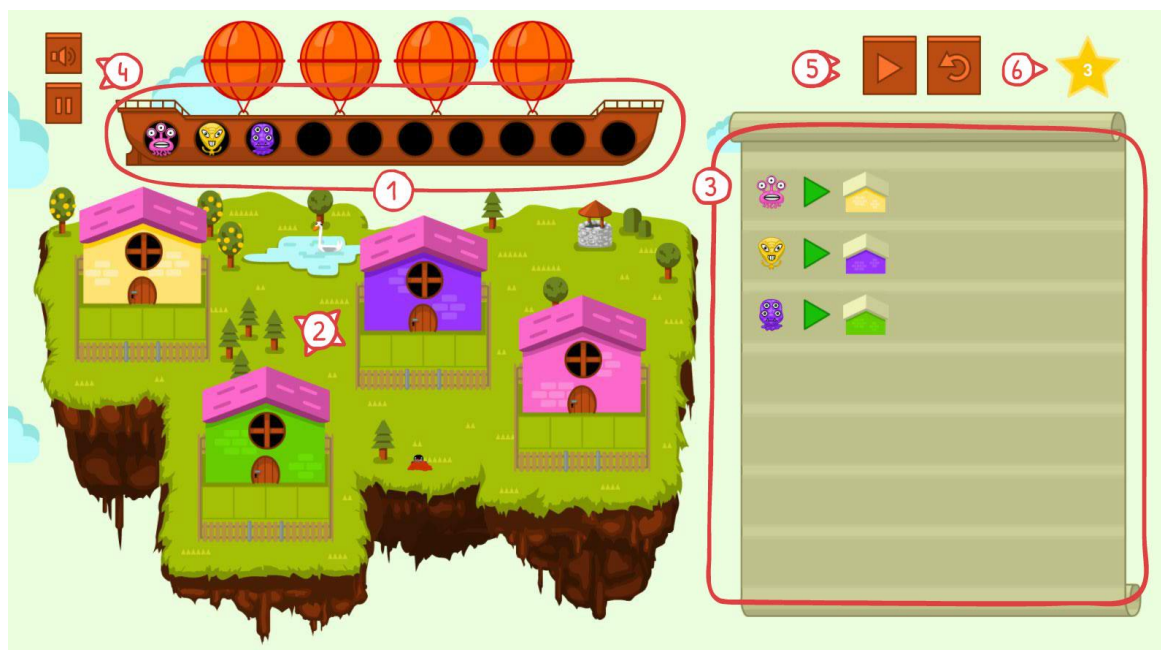
Herní příběh

Na planetě, kam hráč přilétá, havarovala skupina mimozemšťanů v létajícím talíři. Místní starosta Elliot je naštěstí lidumil a nerad by nechal mimozemské návštěvníky bez domova. Než se jejich UFO opraví, je třeba pro mimozemšťany najít náhradní bydlení. Úkolem hráče je rozmístit mimozemšťany do domečků podle grafického návodu od starosty Elliota.

Struktura a jednotlivé prvky

V této sekci je popsána aktuálně finální 5. verze testu *Elliot*. Test je složen z 23 položek, z nichž 19 je vyhodnocovaných pro výpočet standardního skóru. První čtyři velmi snadné položky slouží jako zácvik.

Ukázka druhé zácvičné položky je uvedena níže na Obrázku 4. V každé položce je úkolem hráče ubytovat mimozemšťany do domečků dle graficky znázorněných podmínek. Každou položku je přitom možné pomyslně rozdělit na šest herních segmentů, které jsou označeny na Obrázku 4. V horní části obrazovky jsou umístěni mimozemšťané, které je třeba ubytovat (1). V hlavní části herní obrazovky jsou rozmístěny prázdné domečky k zaplnění (2). Vpravo se nachází návod v podobě pergamentu, na kterém jsou uvedeny graficky znázorněné podmínky tvořené piktogramy (3). Vlevo nahoře jsou umístěny ovládací prvky společné pro všechny hry systému Invenio – pauza a možnost ztlumit nedůležité zvuky (4). Nad návodem (5) jsou umístěny další prvky zásadní pro ovládání hry – tlačítka play a reset, jejichž funkce je popsána níže v textu. Vedle nich je umístěna hvězdička signalizující počet získaných bodů (6).



Obrázek 4. Příklad položky testu *Elliot* se znázorněním jednotlivých prvků.

Základními prvky testu jsou mimozemšťané a domečky. Každý mimozemšťan má vždy právě dvě charakteristiky – barvu (růžová, fialová, zelená a žlutá) a počet očí (1-4). Na počátku hry mají domečky jen jednu charakteristiku, kterou je barva zdi (varianty jsou totožné s barvami mimozemšťanů). Počet charakteristik domečků se v průběhu hry postupně zvyšuje až na 4 (viz dále).

Základním pravidlem je, že žádný mimozemšťan nesmí zůstat nebytovaný. V úvodní části testu také platí, že do každého domečku se vejde právě jeden mimozemšťan. Toto pravidlo se však v průběhu testu změní (viz dále).

Podmínky v návodu (pergamenu) jsou dvojího typu – jednoduché a složené. Jednoduchá podmínka se skládá ze tří piktogramů – první piktogram vždy označuje mimozemšťana či skupinu mimozemšťanů, druhý piktogram představuje relaci (kladnou či zápornou) mezi prvním a třetím piktogramem a třetí piktogram odkazuje obvykle na charakteristiku domečku (případně v pozdější fázi hry na konkrétního mimozemšťana či skupinu mimozemšťanů). Příkladem jednoduché podmínky může být např. zápis z Obrázku 5, který je možné číst takto: „Zelený mimozemšťan s jedním okem (1. piktogram) bydlí (2. piktogram) v některém domečku s růžovou zdí (3. piktogram).“.



Obrázek 5. Ukázka jednoduché podmínky z návodu.

U obtížnějších položek se objevují složené podmínky, které jsou tvořeny dvěma jednoduchými podmínkami spojenými piktogramem s relačním významem. Složená podmínka je tedy celkem tvořena 7 piktogramy (3 za každou z jednoduchých podmínek a 1 relačním piktogramem mezi nimi). Příkladem složené podmínky může být např. zápis z Obrázku 6, který je možné číst takto: „Zelený mimozemšťan s jedním okem bydlí v některém domečku s růžovou zdí a zároveň zelený mimozemšťan s jedním okem nebydlí v domečku s fialovou střechou.“.



Obrázek 6. Ukázka složené podmínky z návodu.





Položky jsou seřazeny tak, aby se zvyšovala jejich obtížnost, tedy náročnost logických operací nutných k vyřešení položky. Současně se od sebe položky liší svou komplexitou. Komplexita položek je přitom vyjádřena množstvím použitých prvků (mimozemšťanů, domečků a podmínek) a testových principů. Každá položka obsahuje:

- tři až šest domečků (nejčastěji čtyři),
- jednoho až osm mimozemšťanů (nejčastěji čtyři)
- jednu až pět podmínek (ať už jednoduchých, či složených; nejčastěji tři).

V následující Tabulce 2 jsou popsány jednotlivé testové principy, se kterými je hráč postupně (v pořadí dle tabulky) seznamován. Součástí představení nového testového principu je obvykle také zavedení nového piktogramu, který je využíván pro znázornění podmínek.

Tabulka 2. Popis jednotlivých testových principů.

Herní princip	Ukázka piktogramu	Popis
Mimozemšťan dané barvy a počtu očí		Tři základní typy piktogramů, které označují konkrétní mimozemšťany dané barvy a počtu očí, domečky dle barvy zdi a kladnou relaci mezi nimi, kterou je možné číst jako „bydlí“.
Domečky se zdi dané barvy		
Kladná relace		
Domečky se střechou dané barvy		Nová charakteristika domečku – barva střechy s příslušnými piktogramy.
Záporná relace		Nový piktogram relace „nebydlí“.
Konjunkce		Nový piktogram relace „a zároveň“, který je používán ve složených podmínkách pro spojení dvou jednoduchých podmínek týkajících se stejného mimozemšťana.
Všichni mimozemšťané dané barvy		Nový piktogram označující všechny mimozemšťany dané barvy.
Všichni mimozemšťané s daným počtem očí		Nový piktogram označující všechny mimozemšťany s daným počtem očí.
Více mimozemšťanů v jednom domečku	NA	Nová charakteristika domečku – různý počet okýnek. Odteď mohou být v jednom domečku ubytování v závislosti na počtu okýnek až tři mimozemšťané. Tento herní princip nepředstavuje nový piktogram, ale využívá již zavedené piktogramy konkrétních mimozemšťanů či skupin mimozemšťanů pro znázornění spolubydlení. Příkladem zápisu může být věta: „Tento mimozemšťan bydlí s mimozemšťanem s tímto počtem očí a dané barvy.“

Cokoliv, co má danou barvu		Nový piktogram, který označuje vše, co má danou barvu – mimozemšťany, střechu a zeď domečků. Příkladem zápisu může být věta: „Tento mimozemšťan je spojen s něčím, co má zelenou barvu.“
Zvířátka		Nová charakteristika domečku – zvířátka. U každého domečku mohou být přítomna až 4 zvířátka – pes, kočka, slepice a prase. Pro každé zvíře je zaveden nový piktogram.
Exkluzivní disjunkce		Nový piktogram, který je využíván ve složených podmínkách jako relace mezi dvěma jednoduchými podmínkami – „buď platí první podmínka, nebo platí druhá podmínka“.
Ekvivalence		Nový piktogram, který je využíván ve složených podmínkách jako relace mezi dvěma jednoduchými podmínkami – „buď obě podmínky platí, nebo obě podmínky neplatí“.

Pozn.: Principy jsou uvedeny dle pořadí, ve kterém se objevují v testu.

Každý herní princip je hráči vysvětlen pomocí instrukce, která je kombinací mluveného slova a grafického znázornění (jednoduchých animací a obrázků). Celkem je hra tvořena 12 takovými instrukcemi. První tři herní principy (viz Tabulku 2) jsou představeny v rámci první instrukce. Druhá instrukce znovu opakuje a zdůrazňuje základní pravidlo, že je nutné vždy ubytovat všechny mimozemšťany. Každá další instrukce pak představuje a vysvětluje nový testový princip.

Hráč test ovládá přesouváním mimozemšťanů po herní ploše. Mimozemšťany je přitom možné umístit přímo do domečku, a nebo také provizorně kamkoli na herní plochu. Během vývoje testu se totiž ukázalo, že některé děti tuto možnost využívají k redukcí nároků na pracovní paměť a zachycení průběžného řešení (např. ve chvíli, kdy se rozhodují mezi dvěma sousedními domečky, a k vyřešení potřebují informace z dalších souvisejících podmínek). Tato možnost jim tedy byla ve finální verzi testu ponechána. Kromě tohoto prvku byla pro snížení vlivu paměti do hry zavedena možnost kdykoliv kliknout na kterýkoli piktogram z návodu a zobrazit tak jeho význam (formou tabulky s textem).

Během hry je možné kliknutím na tlačítko „reset“ vrátit úlohu do výchozího stavu, tedy zrušit dosavadní rozmístění mimozemšťanů. K další položce je možné se přesunout pomocí tlačítka „play“. Po kliknutí na něj dostává hráč zpětnou vazbu, zda byla jeho odpověď správná či chybná. Každá položka má právě jedno správné rozmístění mimozemšťanů a na řešení není stanoven časový limit.

Jedním z herních prvků, který se objevuje ve všech testech systému Invenio, je sbírání bodů. V případě metody *Elliot* získává hráč za každou správnou odpověď 3 body, v případě chyby 1 bod. Cílem tohoto prvku je zvýšit motivaci a zájem o test. Nepředpokládáme přitom, že má tento herní prvek vliv na samotné testování.

Verze 1

V období od listopadu do prosince 2016 do června 2017 vznikala a byla ověřována prvotní papírová verze testu s provizorní grafikou.

Základní koncepce testu vznikala na bázi výše popsaných východisek (viz sekci *Východiska a zaměření*). Vzhledem k tomu, že metoda *Elliot* není primárně inspirována principy běžně používaných testů, byla těmito východisky především odborná literatura zaměřená na teorii C-H-C a fluidní inteligenci a výzkumné studie týkající se deduktivního usuzování a dalších souvisejících konstruktů. Celý koncept testu a jeho jednotlivé principy tak vznikaly nově a bylo třeba je ověřit. Z tohoto důvodu vznikla prvotní verze testu nejdříve v papírové podobě s podněty v provizorním grafickém zpracování.

Základní princip testu byl totožný s finální verzí metody, která je detailně vysvětlena výše v sekci *Popis a struktura*. Odlišnosti spočívaly v některých testových principech a konkrétních položkách. Tato pilotní verze 1 byla individuálně administrována v průběhu ledna až června 2017 celkem 15 dětem různého

věku (předškoláci až žáci 7. ročníku ZŠ) třemi administrátory. Data z těchto kognitivních pilotáží byla průběžně kvalitativně zpracovávána a docházelo k výrazným úpravám instrukcí, redukci a naopak tvorbě nových dílčích testových principů, vývoji nových a úpravě stávajících položek apod. Celkem bylo v této fázi vývoje ověřováno více než 60 položek.

Verze 2

V průběhu června a července 2017 vznikala a byla ověřována druhá (elektronická) verze testu, taktéž s provizorní grafikou.

Na základě výsledků ověřování verze 1 byly vybrány nejvhodnější testové principy (shodné s finální verzí metody, viz Tabulku 2) a 29 nejlépe fungujících položek a test byl (ve stejné provizorní grafické podobě) převeden do elektronické formy za použití softwaru PsychoPy (Peirce, 2007). Následně byl test v průběhu června 2017 na přenosných noteboocích individuálně administrován 46 probandům (11 předškolákům, 21 žákům 1. třídy, 14 žákům 2. třídy) v MŠ Proškovo náměstí (Brno) a ZŠ Brno, náměstí Svornosti. Společně byly s touto verzí testu administrovány Ravenovy progresivní matice – barevné (Raven et al., 1991) k orientačnímu ověření souběžné validity. Po sérii statistických analýz, která zahrnovala položkovou analýzu, screening chybějících dat, odhad reliability, analýzu časů řešení jednotlivých položek a korelaci s Ravenovými maticemi bylo rozhodnuto pokračovat ve vývoji metody.

Verze 3

V období od července 2017 do února 2018 vznikala a byla ověřována třetí pilotní verze testu (jednalo se o elektronickou verzi s již finální grafikou). Na všech fázích tohoto procesu (včetně přípravy standardizační verze) se v rámci své studie podílela Pálková (2018)⁶.

Tvorba

Po ověření verze 2 byl test nově naprogramován v rámci systému Invenio a jednotlivé prvky metody byly převedeny do stávající grafické podoby. Tato verze se od finální verze 5 popsané v sekci *Struktura a jednotlivé prvky* lišila jen nepatrně, a to především ve formě instrukcí. V této verzi testu měly instrukce podobu videí s komentářem, ve kterých byly jednotlivé testové principy představovány na konkrétních položkách s názornou ukázkou jejich řešení. Ve finální verzi metody je většina testových principů a nových piktogramů představována obecně, bez návaznosti na konkrétní položku.

Na základě analýz z ověřování verze 2 byly dále upraveny jednotlivé položky a současně došlo k tvorbě řady nových. Kombinací původních, upravených a nových položek vznikly dva pilotní sety úkolů (Set 1 a 2). Každý z těchto setů se skládal z 22 položek (z toho první 4 položky byly zácvičné a totožné pro oba sety). Kromě těchto základních setů byl připraven také tzv. doplňkový set s 9 položkami a novým testovým principem (Set 3). U položek, u kterých byl tento princip použit, musel hráč posoudit, zda je možné na základě prezentovaných podmínek položku vyřešit správně. Pokud ano, položka se dále řešila standardním způsobem (rozmíst'ování mimozemšťanů do domečků), pokud ne, hráč klikl na nové tlačítko „nelze vyřešit“.

Kognitivní pilotáž

V lednu a únoru 2018 byla provedena kognitivní pilotáž této verze na 8 žácích 2. – 6. tříd ZŠ. V rámci individuální administrace testu bylo nejdříve ověřováno, zda děti nemají problém s rozlišováním použitých barev (především růžové a fialové). V tomto případě se neobjevily žádné problémy. Současně bylo ověřováno jednoznačné pochopení jednotlivých instrukcí, ovládnání nové verze testu a délka administrace.

Na základě pozorování z kognitivní pilotáže došlo k drobným úpravám instrukcí, byla přidána možnost zobrazit si kliknutím význam jednotlivých piktogramů v nápovědě a upravena grafická podoba

⁶ Ve studii Pálkové (2018) je 3. pilotní verze testu *Elliot* označována pracovním názvem *E. T. Rover*.

některých piktogramů tak, aby byly lépe odlišitelné mezi sebou a jejich význam byl pro děti jednoznačnější.

Sběr dat a vzorek

Sběr dat pro ověřování verze 3 probíhal v únoru 2018. Test byl celkem zadán 227 žákům (z toho 44 % bylo dívek) 1.-5. tříd ZŠ. Velikost vzorku pro jednotlivé sety je uvedena v Tabulce 3.

Tabulka 3. Velikost vzorku pro jednotlivé sety použité při ověřování Pilotu 3.

Set	Celkové <i>n</i>	1. třída	2. třída	3. třída	4. třída	5. třída
Set 1	114	15	24	18	26	31
Set 2	113	14	26	20	25	28
Set 3	125	9	22	25	32	35

Testování žáci pocházeli ze 2 základních škol – ZŠ a MŠ Ždírec nad Doubravou a ZŠ Havlíčkův Brod, Nuselská. Zapojení do studie bylo dobrovolné, anonymní a vzhledem k nezletilosti žáků probíhalo pouze na základě písemného souhlasu zákonného zástupce.

Společně s metodou *Elliot* byl zadán *Cattellův test fluidní inteligence (CFT 20-R)*; Fajmonová et al., 2015) pro ověření souběžné validity (viz sekci *Souběžná a diferenciatní validita* níže, kde je také detailněji popsána metoda *CFT 20-R*). Obě metody administrovala členka výzkumného týmu Invenio (u menších dětí ve spolupráci s asistenty pedagoga). V případě testu *Elliot* byli žáci náhodně rozřazeni k jednomu ze dvou setů položek. Doplnkový set byl administrován všem dětem, které stihly vyřešit všechny úlohy hlavního setu před koncem hodiny.

Výsledky

Následně byla realizována položková analýza. Fungování všech položek testu bylo ověřováno pomocí tří kritérií – popularity (podílu správných odpovědí pro jednotlivé ročníky), diskriminační účinnosti (korigovaná korelace položky s celkovým skóre) a korelace se standardním skóre testu *CFT 20-R*. Faktorová struktura obou hlavních setů (Set 1 a 2) byla ověřována pomocí explorační faktorové analýzy na matici polychorických korelací.

Na základě výsledků položkové analýzy nebyly objeveny žádné problémy týkající se testových principů obou hlavních setů. Oba byly jednodimenzionální, vykazovaly dostatečnou míru reliability (ve formě vnitřní konzistence) a dle očekávání také středně silné korelace se skóre testu *CFT 20-R*. Ověřován byl také nový testový princip zařazený do doplnkového setu (Set 3), který vyžadoval po hráči posouzení, zda je možné na základě prezentovaných podmínek dojít k jednoznačně správnému řešení položky. V rámci analýz se však ukázalo, že tento testový princip není dostatečně funkční, a nebyl proto zařazen do finální verze testu.

Na základě výsledků analýz bylo vybráno 25 nejlépe fungujících položek, které byly zařazeny do verze 4, která byla standardizována. Detailní zdůvodnění výběru položek je uvedeno ve studii Pálkové (2018). Dále byly upraveny instruktážní videa testu tak, aby představovaly jednotlivé testové principy bez návaznosti na konkrétní položky.

Standardizace

Tato sekce popisuje standardizační studii 4. verze testu *Elliot*. Analýzy byly prováděny v paradigmatech klasické testové teorie (CTT) a teorie odpovědi na položku (IRT). Reliabilita testu, normy a validita jsou diskutovány v následujících samostatných sekcích.

Sběr dat a popis vzorku

Test *Elliot* byl standardizován na vzorku 1011 žáků (1. až 6. ročníku) 16 českých základních škol. Stručný popis vzorku je k dispozici v Tabulce 4.

Tabulka 4. Popisné statistiky standardizačního vzorku.

Ročník	N	Typický věk	% dívek
1	89	7	42
2	166	8	55
3	182	9	53
4	242	10	50
5	150	11	47
6	182	12	48

Sběr dat probíhal od října do prosince 2018. Účast byla dobrovolná, vyžadující písemný informovaný souhlas zákonného zástupce. Hru administrovali vyškolení učitelé či školní psychologové příslušných škol, a to v počítačových učebnách.

Školy podílející se na standardizaci jsou uvedeny v Tabulce 5. Vybrány byly příležitostně na základě ochoty podílet se na studii. Potenciálně participující instituce však byly voleny tak, aby pokrývaly velikosti sídel typické pro Českou republiku. Rozdělení participantů dle velikosti sídla je k dispozici v Tabulce 6.

Tabulka 5. Školy podílející se na standardizaci.

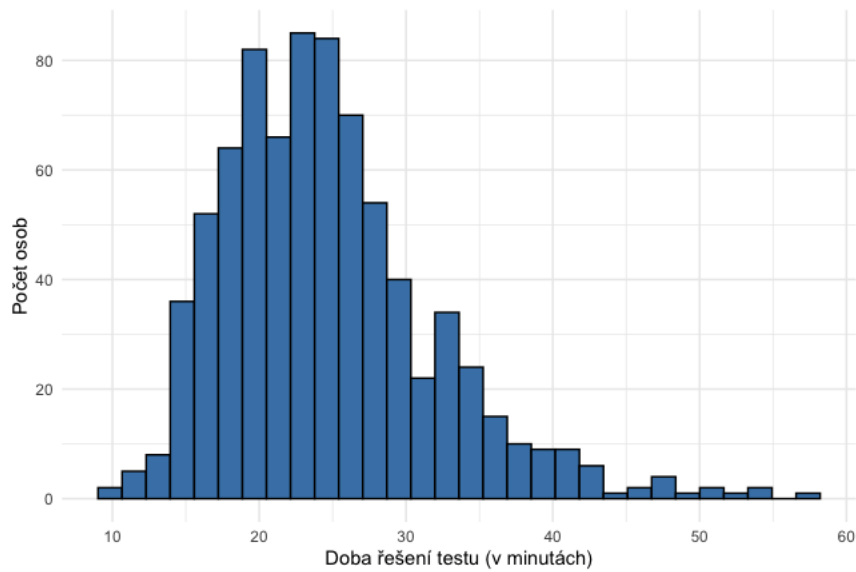
Název školy	Kraj
ZŠ a MŠ Jana Broskvy	Jihomoravský
ZŠ a MŠ Benešov nad Ploučnicí	Ústí nad Labem
ZŠ Humpolec	Vysočina
ZŠ a MŠ Frenštát pod Radhoštěm	Moravskoslezský
ZŠ Malšovice	Ústí nad Labem
ZŠ nám. Curieových, Praha	Praha
ZŠ a MŠ Kukleny	Královehradecký
ZŠ Mládežnická, Havířov	Moravskoslezský
ZŠ Litomyšl	Pardubický
ZŠ Horní Slavkov	Karlovarský
ZŠ Kovářov	Jihočeský
ZŠ a MŠ Norská, Kladno	Středočeský
ZŠ Jedovnice	Jihomoravský
ZŠ Větrná, Uherské Hradiště	Zlínský
ZŠ a MŠ Vlasatice	Jihomoravský

Tabulka 6. Rozdělení participantů dle velikosti sídla, kde se nachází jejich školy.

Velikost sídla	Méně než 1 999	2 000 až 9 999	10 000 až 49 999	50 000 a více	Praha
Počet osob	93	112	602	204	14

Odpovědi 14 participantů byly odstraněny pro podezření z nepochopení testových instrukcí. Jinak byla využita data všech účastníků, a to i 220 žáků, kteří test nedokončili.

Mediánový čas řešení celého testu činil 23,7 minut. Rozložení je prezentováno níže na Obrázku 7.



Obrázek 7. Rozložení času řešení celého testu.

Položková analýza

Položková analýza popisuje statistické vlastnosti testových položek. V Tabulce 7 jsou uvedeny popularity (podíly správných odpovědí ve vzorku) po třídách, medián délky řešení v sekundách a korigované korelace položek s celkovým skórem (r_{cor}). První čtyři položky nejsou skórovány, protože slouží jen k zácvicí. Položky označené křížkem byly po standardizaci odstraněny (viz dále).

Tabulka 7. Položková analýza.

Pořadí položky	popularita						čas řešení (Me)	r_cor
	1. ročník (n = 89)	2. ročník (n = 166)	3. ročník (n = 182)	4. ročník (n = 242)	5. ročník (n = 150)	6. ročník (n = 182)		
1*	0,84	0,91	0,94	0,97	0,97	0,97	15,5	0,22
2*	0,91	0,92	0,95	0,98	0,97	0,98	22	0,21
3*	0,96	0,89	0,94	0,96	0,98	0,96	26,5	0,25
4*	0,73	0,87	0,84	0,94	0,98	0,98	26,5	0,35
5	0,81	0,83	0,88	0,94	0,96	0,95	26	0,29
6	0,65	0,77	0,75	0,92	0,91	0,88	38	0,37
7	0,45	0,63	0,65	0,83	0,94	0,86	25,5	0,51
8+	0,47	0,58	0,62	0,66	0,79	0,82	27	0,29
9	0,16	0,40	0,38	0,52	0,68	0,63	38,5	0,50
10	0,31	0,53	0,55	0,70	0,79	0,76	45,5	0,47
11	0,21	0,45	0,47	0,66	0,74	0,74	58	0,54
12	0,11	0,17	0,29	0,39	0,47	0,45	49,5	0,46
13	0,16	0,20	0,26	0,41	0,47	0,43	65,5	0,46
14	0,07	0,16	0,26	0,37	0,45	0,45	35,5	0,51
15	0,05	0,14	0,27	0,33	0,47	0,45	59,5	0,56
16	0,15	0,41	0,55	0,68	0,77	0,74	78,5	0,56
17	0,10	0,23	0,27	0,43	0,46	0,44	83	0,45
18	0,01	0,15	0,14	0,39	0,48	0,47	55,5	0,52
19	0,06	0,07	0,19	0,30	0,39	0,35	71,5	0,45
20	0,02	0,06	0,10	0,18	0,22	0,27	64,5	0,36
21	0,09	0,19	0,21	0,39	0,44	0,35	73	0,46
22+	0,09	0,11	0,13	0,16	0,18	0,21	71	0,18
23	0,13	0,24	0,35	0,49	0,53	0,47	102,5	0,42
24	0,03	0,06	0,07	0,09	0,12	0,11	78	0,24
25	0,01	0,06	0,11	0,19	0,22	0,22	84,5	0,52

* = položka není skórována

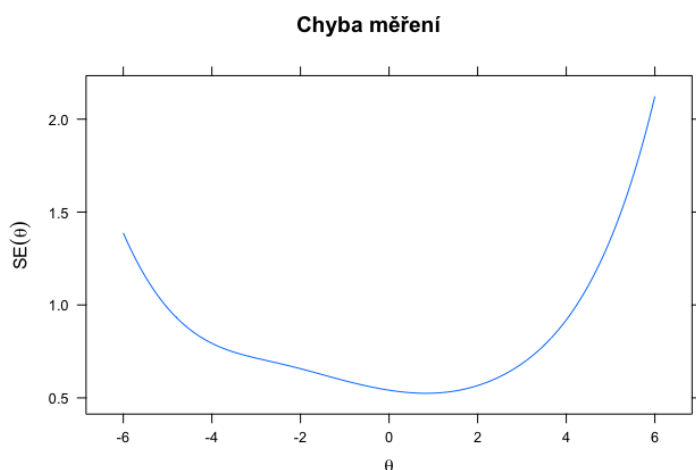
+ = položka byla po standardizaci vyřazena

IRT modely

Standardizační data byla nejprve analyzována pomocí Raschova (Bond et al., 2020) modelu pro všechny participanty společně. Do modelu vstupovaly všechny položky až na první čtyři (1 – 4), které plní pouze závěšnou funkci a nejsou skórovány. Model byl odhadnut pomocí balíčku *mirt* (1.32.1) v prostředí R (R Core Team, 2020). Raschův model předpokládá jednodimenzionální charakter položek.

Shoda modelu s daty byla adekvátní (RMSEA = 0,027 ₉₀CI [0,021 ; 0,032], TLI = 0,979, SRMSR = 0,052).

Křivka chyby měření Raschova modelu (znázorněna níže na Obrázku 8) reprezentuje chybu měření v závislosti na schopnosti participanta. Schopnost participanta je latentním skórem, pro nějž je v rámci Raschova modelu předpokládáno normální rozložení. Toto rozložení má v případě jednoskupinového modelu průměr 0 a modelem odhadovanou směrodatnou odchylku, v tomto případě $s = 1,4$. Pro hodnoty od 0 do $|2,8|$ na ose x, což přibližně odpovídá ± 2 SD latentního skóru, se chyba měření pohybuje od 0,5 do 0,7; nad těmito hodnotami začíná strmě růst, pro hodnoty latentního skóru od -4 do 4 však leží pod 1.



Obrázek 8. Křivka chyby měření (inverzní k lokální reliabilitě z Obrázku 8).

Indexy položkového fitu (*infit* a *outfit*) popisující úroveň, s jakou se každá z položek shoduje s modelem, jsou uvedeny v Tabulce 8. Za nepřijatelné jsou obecně považovány položky s hodnotou *infit* a *outfit* větší než 2 (Linacre, 2002).

Tabulka 8. Infit a outfit položek (Raschův model)

Pořadí položky	Outfit	Infit
1*	0,83	0,97
2*	1,15	0,99
3*	1,03	1,03
4*	0,83	0,96
5	1,31	1,01
6	1,02	1,01
7	0,81	0,92
8+	1,56	1,21
9	0,99	0,97
10	1,12	1,00
11	0,89	0,95
12	0,93	1,00
13	1,15	1,04
14	0,97	1,00
15	0,83	0,94
16	0,95	0,96
17	1,02	1,04
18	0,87	0,96
19	1,02	1,04
20	1,17	1,06
21	1,06	1,03
22+	1,76	1,11
23	1,29	1,10
24	1,35	1,05
25	0,75	0,88

* = položka není skórována

+ = položka byla po standardizaci vyřazena

Na základě této a předchozí položkové analýzy (viz Tabulku 7) byly z finální verze testu odstraněny položky č. 8 a 22. Důvodem pro jejich vyřazení byla relativně nižší hodnota korigované korelace s celkovým skórem a relativně horší hodnota *oufit*. Další informace o výsledné sadě položek obsahuje sekce *Struktura a jednotlivé prvky*.

Reliabilita

Reliabilita byla odhadována pomocí dvou metod – split-half (klasická testová teorie) a IRT.

Split-half Reliabilita

Split-half reliabilita je metodou výpočtu reliability zohledňující narůstající obtížnost položek. Počítá se jako korelace počtu správných odpovědí na liché a sudé položky. Split-half reliabilita testu *Elliot* odpovídá $r = 0,77$, což ukazuje na spolehlivost odpovídající účelu testu. Tabulka 9 shrnuje hodnoty split-half reliability pro jednotlivé ročníky. Je zřejmé, že test dosahuje ve všech ročnících spíše nízkých hodnot split-half reliability, pro první ročník je však hodnota reliability zjevně nejvyšší. Pro zbylé ročníky lze hodnoty reliability považovat vzhledem ke skříninkovému účelu testu za relativně uspokojivé.

Tabulka 9. Odhad split-half reliability pro jednotlivé ročníky.

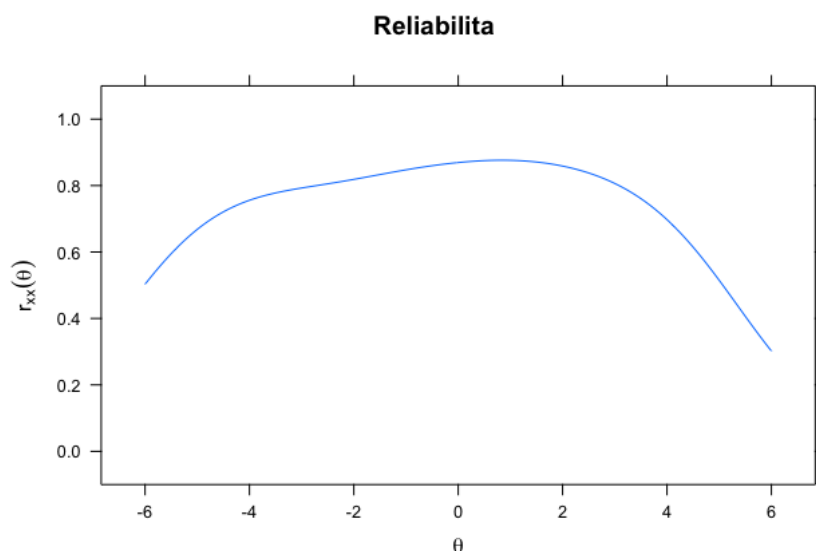
Ročník	Split-half reliabilita
1	0,58
2	0,70
3	0,74
4	0,69
5	0,73
6	0,69

IRT reliabilita

Odhad IRT reliability je založen na Raschově modelu aplikovaném na sebraná data. Takto odhadnutá marginální reliabilita je podobná odhadu pomocí split-half (0,76; empirická reliabilita: 0,84⁷).

Na Obrázku 9 je uveden odhad lokální reliability testu. IRT přístup umožňuje odhadovat reliabilitu testu v závislosti na schopnosti účastníka (tj. brát v úvahu chyby měření lišící se pro různé úrovně latentní schopnosti). Průměrné schopnosti a jejich odchylky pro jednotlivé ročníky jsou k dispozici v sekci *Normy*.

⁷ Pro popis rozdílu mezi marginální a empirickou reliabilitou z balíčku *mirt* pro jazyk R viz např. <https://stats.stackexchange.com/questions/427631/difference-between-empirical-and-marginal-reliability-of-an-irt-model>



Obrázek 9. Lokální reliabilita testu získaná z Raschova modelu. Řecké písmeno theta symbolizuje latentní schopnost.

Obrázek 9 demonstruje přijatelnou reliabilitu testu (0,7 až 0,8) v rozpětí schopností $[-2,8 ; 2,8]$, což zhruba odpovídá ± 2 SD latentního skóru. Extrémně schopní participanti jsou měřeni se sníženou spolehlivostí. Test si nicméně zachovává uspokojivý standard pro většinu nadaných žáků ($> 0,7$), byť reliabilita u mimořádně nadprůměrných žáků (3 SD nad průměrem) klesá do pásma $[0,7 - 0,4]$.

Skupinový model

Pro odhady průměrů distribucí latentních schopností a jejich směrodatných odchylek pro jednotlivé ročníky jsme vycházeli ze skupinového Raschova modelu s ročníky jako skupinami. Obtížnosti položek byly napříč skupinami omezeny na stejnou hodnotu, průměr latentní distribuce pro první ročník byl omezen na hodnotu 0. Průměry latentních distribucí ostatních ročníků a směrodatné odchylky všech latentních distribucí byly pak volně odhadovány. V modelu nebyly použity zácvičné ani vyřazené položky (viz Tabulka 7).

Skupinový model se všemi ročníky (1 – 6) jako skupinami vykazoval nedostatečnou shodu s daty (RMSEA = 0,019 $_{90}$ CI [0,017 ; 0,022], SRMSR (1. ročník) = 0,226, SRMSR (2. ročník) = 0,97, SRMSR (3. ročník) = 0,109, SRMSR (4. ročník) = 0,084, SRMSR (5. ročník) = 0,114, SRMSR (6. ročník) = 0,105, TLI = 0,90). Zejména u 1. je velikost průměrných reziduí příliš vysoká; indexy TLI a RMSEA jsou adekvátní. Jestliže do modelu zahrneme pouze 2. - 6. ročníky, shoda modelu s daty se očekávaně zlepší (RMSEA = 0,015 $_{90}$ CI [0,009 ; 0,020], SRMSR (2. ročník) = 0,117, SRMSR (3. ročník) = 0,108, SRMSR (4. ročník) = 0,078, SRMSR (5. ročník) = 0,097, SRMSR (6. ročník) = 0,100, TLI = 0,96).

Výše uvedená zjištění ve spojení s nízkou reliabilitou u prvního ročníku nás vedla k závěru, že test je vhodný pouze pro žáky druhých až šestých ročníků.

Odhadnuté hodnoty průměrů a směrodatných odchylek latentních skóru na základě skupinového Raschova modelu pro 2. – 6. ročníky jsou uvedeny v Tabulce 10. Pro odhad normativního modelu byly ze vzorku odstraněny děti s diagnostikovaným mimořádným nadáním ($n = 14$), aby vzhledem ke svému nerovnoměrnému zastoupení v ročnících nezkreslovaly ročníkové výsledky.

Tabulka 10. Popisné statistiky latentních skóru dle ročníku.

Ročník	Průměrný skór	SD skóru
2	0	1,2
3	0	1,3
4	0,7	1,2
5	1,1	1,1
6	1,1	1,0

Pozn.: Průměrné skóry jsou fixovány na hodnoty větší nebo rovno nule.

Validita

V této sekci představujeme důkazy pro naše tvrzení, že test *Elliot* je validním nástrojem pro měření *Fluidní inteligence* a *Deduktivního usuzování*.

Obsahová validita

Jak je uvedeno v kapitole *Východiska a zaměření*, test *Elliot* byl vyvíjen primárně na základě C-H-C teorie inteligence (např. Schneider & McGrew, 2018) a v souladu s již ověřenými testovými principy.

Souběžná a diferenciální validita

Souběžnost Pilotní verze 3 s CFT 20-R

Pálková (2018) posuzovala souvislost verze 3 s testem *CFT 20-R* (Fajmonová et al., 2015). *CFT 20-R* je neverbální test *Fluidní inteligence* zaměřený primárně na *Induktivní usuzování*. Je administrovaný skupinově dětem ve věku od 7,5 do 15 let. Test sestává ze dvou setů, z nichž každý má čtyři subtesty. Administrovat je možné pouze první set či oba. První set je také možné použít se standardním, či s prodlouženým časovým limitem. V případě studie Pálkové (2018) byl administrován pouze první set a to s prodlouženým časovým limitem. Vnitřní konzistence celého testu je odhadována na úrovni Cronbachovy alfy = 0,88.

Studie byla realizována na vzorku žáků druhých až šestých tříd; žákům prvních tříd nebyl test *CFT 20-R* vzhledem k jeho formě a věkovému limitu administrován. Podrobnosti o vzorku této studie jsou k dispozici v Tabulce 3 v sekci *Verze 3*. Test *Elliot* byl administrován skupinově v počítačových učebnách spolupracujících škol; *CFT 20-R* byl skupinově zadáván tamtéž, avšak v učebnách standardních. Mezi administracemi byla alespoň hodinová pauza. Většinou však sběr dat probíhal v různé dny. Nejdelší doba mezi administracemi činila 2 dny.

Pearsonův korelační koeficient odhadu latentních skóru v testu *Elliot* a standardních skóru z *CFT 20-R* činil $r = 0,64$ ($p < 0,01$, $N = 99$) pro první set a $r = 0,54$ ($p < 0,01$, $N = 99$) pro druhý. Výsledek lze vyložit jako důkaz validity testu *Elliot*, protože vykazuje souběžné výsledky se široce užívaným testem *Fluidní inteligence*. Korelace je vyšší, než bylo očekáváno, a to z důvodu rozdílů mezi oběma testy – zejména v jejich formě (*Elliot* je gamifikován a administrován na počítači bez časového limitu, zatímco *CFT 20-R* je klasickým testem papír a tužka se striktním časovým omezením) a jejich zaměření (*CFT 20-R* cílí na úzkou schopnost indukce, zatímco *Elliot* je zaměřen na *Deduktivní usuzování*). Odpovídající testy v původní americké verzi baterie *Woodcock-Johnson IV (WJ IV; McGrew et al., 2014) – Analýza-Syntéza* (podobný testu *Elliot*) a *Formování konceptů* (podobný metodě *CFT 20-R*) – korelují $r = 0,48$ (pro šesti- až osmileté) a $r = 0,52$ (pro devíti- až třináctileté).

Souběžnost standardizační verze a CFT 20-R

Souvislost mezi *CFT 20-R* (jeho prvním setem s prodlouženým časovým limitem) a 5. verzí testu *Elliot* ověřovala také Wranová (2020) na vzorku 154 žáků třetích tříd. Testy byly administrovány skupinově, s jednodenním rozestupem. Studie se zúčastnilo 6 škol – ZŠ Provoznická, Ostrava, ZŠ Vrchní, Opava, ZŠ a MŠ Bolatice, ZŠ Štěpánkovice, ZŠ T.G. Masaryka, Opava, ZŠ a MŠ MUDr. Emílie Lukášové, Ostrava.

Pearsonův korelační koeficient odhadů latentních skóre odpovídal $r = 0,45$ ($p < 0,01$, $N = 154$). Výsledky naznačují, že testy jsou souběžné natolik, aby poskytly důkazy o validitě testu *Elliot*. Z důvodů uvedených v předchozí sekci se větší překryv obou testů nepředpokládal.

Divergence standardizační verze a Obrácených číselných řad

Ve shora zmíněné studii Wranové (2020) autorka rovněž odhadovala vztah mezi 5. verzí testu *Elliot* a nástrojem pro měření *Krátkodobé pracovní paměti* – testem *Obrácené číselné řady* z baterie *Woodcock-Johnson: International edition II – Tests of Cognitive Abilities (WJ IE II COG*; Ruef et al., 2010). Test je administrován individuálně. Úkolem testované osoby je v opačném pořadí opakovat administrátorem vyřčené číselné řady.

Obrácené číselné řady byly administrovány bezprostředně po testu *Elliot*, maximálně následující den. Pearsonův korelační koeficient odhadu latentních skóre v testu *Elliot* a hrubých skóre z *Obrácených číselných řad* odpovídal $r = 0,18$ ($p < 0,01$, $N = 154$), což ukazuje na slabou pozitivní asociaci mezi oběma nástroji. Tento výsledek odpovídá našemu očekávání – předpokládáme totiž, že se *Krátkodobá pracovní paměť* podílí na řešení úloh primárně zaměřených na *Fluidní inteligenci*, zároveň jde však v rámci C-H-C teorie o oddělenou schopnost. Tuto úvahu lze dále doložit korelací testů *Analýza a syntéza* (test svým zaměřením blízký metodě *Elliot*) a *Obrácené číselné řady* v baterii *WJ IV* (McGrew et al., 2014), která je odhadována na $r = 0,40$ (pro šesti- až osmileté) a $r = 0,34$ (pro devíti- až třináctileté). Je možné, že se v námi zjištěné nižší korelaci odráží testové a ovládací prvky, které jsme do hry zařadili cíleně s úmyslem snižovat vliv *Krátkodobé pracovní paměti* na výkon v testu *Elliot* (viz sekce *Struktura hry a jednotlivé prvky*). Současně lze rovněž očekávat, že se daná korelace oslabí v důsledku odlišných forem obou testů – *Elliot* je gamifikován a administrován skupinově, což pro *Obrácené číselné řady* neplatí.

Specifické populace

Wranová (2020) ve své studii zaměřené na žáky třetích tříd rovněž posuzovala odlišné fungování 5. verze testu *Elliot* u dětí se specifickými poruchami učení. Zaměřila se přitom na souvislost testu *Elliot* a několika metod indikujících specifické poruchy učení, konkrétně využila nástroje *Test rychlého jmenování (RAN)* – *obrázky*, *Elize hlásek* a *Čtení souvislého textu*.

*Test rychlého jmenování (RAN)*⁸ – *obrázky* je metodou z diagnostické baterie *MABEL* (Caravolas et al., 2020). Testovanému jsou ukázány dvě sady obrázků, jež má co nejrychleji pojmenovat. Ukazatelem výkonu je čas a počet chyb. Autoři neuvádí žádné informace o reliabilitě testu.

Elize hlásek je testem z *Baterie diagnostických testů gramotnostních dovedností pro žáky 2. až 5. ročníků ZŠ* (Caravolas & Volín, 2005) určeným k měření fonematického porozumění, které se odráží ve fonematické analýze pseudoslov. Posuzuje se přesnost a rychlost odpovědi respondenta. Test má dvě části – v první z nich má testovaný vynechávat druhou hlásku z pseudoslova, ve druhé pak poslední hlásku. Vnitřní konzistence v případě žáků třetích tříd je odhadována jako Cronbachova alfa = 0,75.

Čtení souvislého textu je test z komplexní baterie zaměřené na diagnostiku čtení a psaní⁹ (Bednářová et al., 2015). Testované osoby mají za úkol číst prezentovaný článek (použit byl text s názvem *Veverka*) a to buď 60, nebo 120 sekund (v této studii byl použit limit 60 sekund). Ukazatelem výkonu je počet správně přečtených slov za daný časový limit. Examinátor může též posuzovat čtenářskou techniku. Autoři v manuálu baterie neuvádí žádné informace o reliabilitě testu.

Výše uvedené testy byly individuálně administrovány bezprostředně po testu *Elliot*, maximálně následující den. Autorka vytvořila regresní model s diagnostickými indikátory specifických poruch

⁸ RAN je zkratkou pro Rapid Automated Naming.

⁹ Celý název baterie zní *Diagnostika schopností a dovedností v oblasti čtení a psaní: Varianta pro pedagogy škol a školní poradenská pracoviště, 3. a 4. ročník*.

učení jako prediktory a hrubými skóry z testu *Elliot* jako závislou proměnnou. Uzavřela, že mezi specifickými poruchami učení a výkonem v testu *Elliot* zřejmě není souvislost.

Uvedené indikátory (a prediktory v popsáném modelu) sestávaly z následujících proměnných (pro účely tohoto manuálu byl model proveden znovu s odhady latentních skóru v testu *Elliot*):

- Počet chyb v *Testu rychlého jmenování (RAN) – obrázky* ($\beta = -0,03, p = 0,75$).
- Standardní skóry rychlosti odpovědi ($\beta = -0,3, p = 0,02$) a přesnosti ($\beta = 0,3, p = < 0,01$) v testu *Elize hlásek*.
- Standardní skóry v testu *Čtení souvislého textu* ($\beta = 0,2, p = 0,08$).

Poděkování

Rádi bychom srdečně poděkovali Barboře Pálkové, která se na vývoji testu podílela prostřednictvím své bakalářské práce. Její nápady i nadšení značně napomohly vývoji metody. Stejně chceme poděkovat také Vendule Wranové za čas a snahu, které investovala do validizace testu *Elliot* v rámci své diplomové práce.

Děkujeme rovněž Magistrátu města Brna za finanční podporu při vývoji metody v rámci projektu MMB “0561/2017 (ID=37926) – Vývoj nových diagnostických nástrojů a metod pro plošnou screeningovou a individuální identifikaci mimořádně nadaných dětí”

Literatura

- Bond, T., Yan, Z., & Heene, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences*. Routledge. <https://doi.org/10.4324/9780429030499>
- Caravolas, M., Mikulajová, M., Defior, S., & Seidlová Málková, G. (2020). *Multilanguage Assessment Battery of Early Literacy*. <https://www.eldel-mabel.net/>
- Caravolas, M., & Volín, J. (2005). *Baterie diagnostických testů gramotnostních dovedností pro žáky 2. až 5. ročníků ZŠ*. Institut pedagogicko-psychologického poradenství ČR.
- Bednářová, J., Baldrmannová, S., Burešová, H., Filípková, K., Klusalová, P., Pařenicová, Z., Svobodová, N., Šmardová, V., Trávníková, A., Zárubová, L., Širůček, J., & Urbánek, T. (2015). *Diagnostika schopností a dovedností v oblasti čtení a psaní: Varianta pro pedagogy škol a školní poradenská pracoviště, 3. a 4. ročník*. Pedagogicko-psychologická poradna Brno.
- Benson, N. F., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Fefer, S. A., & Morgan, G. B. (2019). Test use and assessment practices of school psychologists in the United States: Findings from the 2017 National Survey. *Journal of School Psychology, 72*, 29-48. <https://doi.org/10.1016/j.jsp.2018.12.004>
- Brož, J. (2007). *Zebry, zebry, zebříčky*. Scientia.
- Caparos, S., & Blanchette, I. (2017). Independent effects of relevance and arousal on deductive reasoning. *Cognition and Emotion, 31*(5), 1012-1022. <https://doi.org/10.1080/02699931.2016.1179173>
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511571312>
- Cunningham, D. W. (2016). *Set theory: A first course*. Cambridge University Press. <https://doi.org/10.1017/CBO9781316341346>
- Elliot, C. D. (2007). *Differential ability scales-second edition manual*. Harcourt Assessments.
- Evans, J. St. B. T. (2005). Deductive reasoning. In K. J. Holyoak, & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 169-184). Cambridge University Press.
- Fajmonová, V., Hönigová, S., Urbánek, T., & Širůček, J. (2015). *CFT 20-R – Cattellův test fluidní inteligence*. Hogrefe-Testcentrum.
- Hunt, E. (2011). *Human intelligence*. Cambridge University Press
- Johson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology, 50*, 109-135. <https://doi.org/10.1146/annurev.psych.50.1.109>
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children, Second Edition: Manual*. AGS Publishing.
- Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean. *Rasch Measurement Transactions, 16*(2), 878.
- Matějček, Z., Šturma, J., Vágnerová, M., & Žlab, Z. (1992). *Zkouška čtení*. Psychodiagnostika.
- McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Woodcock Johnson IV: Technical manual*. Riverside Publishing.
- Pálková, B. (2018). *Vývoj a pilotní ověření metody k měření deduktivních schopností u dětí pomocí videohry*. Bakalářská práce. Masarykova univerzita.

- Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1-2), 8-13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Politzer, G., & Bourmaud, G. (2002). Deductive reasoning from uncertain conditionals. *British Journal of Psychology*, 93, 354-381. <https://doi.org/10.1348/000712602760146251>
- Raclavský, J. (2015). *Úvod do logiky: klasická výroková logika*. Masarykova univerzita.
- Raven, J. C., Court, J. H., & Raven, J. (1991). *Ravenove progresívne matice farebné (CPM)*. Psychodiagnostika, s.r.o.
- Reynolds, C. R., & Kamphaus, R. W. (2015). *Reynolds Intellectual Assessment Scales (2nd ed.)*. Psychological Assessment Resources.
- Rodriguez-Moreno, D., & Hirsch, J. (2009). The dynamics of deductive reasoning: An fMRI investigation. *Neuropsychologia*, 47, 949-961. <https://doi.org/10.1016/j.neuropsychologia.2008.08.030>
- Schrank, F. A. (2016). How to interpret the WJ IV COG. In F. A. Schrank, S. L. Decker, & J. M. Garruto (Eds.), *Essentials of WJ IV cognitive abilities assessment* (pp. 144-221). John Wiley & Sons.
- Stangroom, J. (2009). *Einstein's Riddle: Riddles, Paradoxes, and Conundrums to Stretch Your Mind*. Bloomsbury USA.
- Wechsler, D. (2014). *Wechsler Intelligence Scale for Children (5th ed.)*. NCS Pearson.
- Wranová, V. (2020). *Vztah mezi výkonem v gamifikovaném testu fluidní inteligence a deficity některých dovedností ve zpracování jazyka*. Diplomová práce. Masarykova univerzita.