

VÝVOJ DIAGNOSTICKÉHO NÁSTROJE Z POHLEDU PROJEKTU DIS

MICHAL JABŮREK¹

Abstrakt: Článek zachycuje jednotlivé fáze vývoje diagnostického nástroje. Ty popisuje skrze zkušenost jednoho z členů týmu projektu DIS, v rámci kterého bylo během tří let standardizováno 12 diagnostických nástrojů určených pro dětskou poradenskou praxi.

Klíčová slova: Vývoj testu, projekt DIS, diagnostické nástroje, diagnostika dětí

Úvod

Cílem tohoto článku je zprostředkovat tříletou osobní zkušenost jednoho z členů týmu projektu DIS, který se zabýval tvorbou a standardizací diagnostických nástrojů pro dětskou poradenskou praxi, a usnadnit tak postup budoucích podobně orientovaných výzkumných snah. Článek je koncipován jako přehled jednotlivých fází procesu vývoje diagnostického nástroje, které jsou v některých případech doplněny konkrétními projektovými zkušenostmi. Je nutné upozornit na to, že vývoj testu je komplexní proces a některé zde uvedené fáze jsou vzhledem k rozsahu problematiky zmíněny pouze okrajově. Z tohoto důvodu jsou následující postřehy a připomínky vhodné především pro ty, kteří s vývojem testů zatím nemají příliš zkušeností, a chtějí získat základní přehled o tom, co tvorba diagnostického nástroje zahrnuje.

Projekt DIS

Projekt DIS (celým názvem „Diagnostika dětí a žáků se speciálními vzdělávacími potřebami“) probíhal v období od 2. 9. 2010 do 31. 8. 2013, ale jak bývá u projektů zvykem, byl jím jeho řešitelským tým pohlcen dlouho před a dlouho po jeho realizaci. Původním řešitelem projektu byl Institut pedagogicko-psychologického poradenství ČR (IPPP ČR). Ten byl však 1. 7. 2011 sloučen s dalšími dvěma institucemi² do nové organizace s názvem Národní ústav pro vzdělávání, školské poradenské zařízení a zařízení pro další vzdělávání pedagogických pracovníků (NÚV³). Hlavním cílem projektu

¹ Katedra psychologie, Fakulta sociálních studií MU, Joštova 10, 602 00 Brno

² Těmito dvěma organizacemi byl Národní ústav odborného vzdělávání (NÚOV) a Výzkumný ústav pedagogický (VÚP).

³ URL: www.nuv.cz

bylo vytvořit či adaptovat 12 diagnostických nástrojů, které pomohou zlepšit dle analýz řešitele ne zcela optimální stav v oblasti pedagogicko-psychologické diagnostiky v ČR.

Tento projekt byl spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky. Evropské sociální fondy jsou prostředím natolik specifickým, že mohou ovlivňovat vývoj a směřování daného projektu, spoluutvářet jeho výstupy a v některých případech přímo definovat povahu problémů, se kterými se projekt během realizace potýká. Z těchto důvodů je vhodné upozornit na to, že se některá odborná rozhodnutí učiněná v projektu DIS neobešla bez přihlédnutí k případným administrativně-projektovým mantinelům.

Vývoj a adaptace diagnostického nástroje

Vývoj či adaptace testu je v zásadě komplikovanou sérií odborných a koordinačně-administrativních rozhodnutí a snah. I přes důkladné plánování a promýšlení postupu poskytuje komplexita tohoto procesu prostor pro neočekávané problémy o opomenutí. Z těchto důvodů je vhodné se před samotnou realizací seznámit se standardy pro vývoj či adaptaci testů, jejichž cílem je případné nedůslednosti minimalizovat. Základním dokumentem jsou v tomto případě Standardy pro pedagogické a psychologické testování (např. AERA, APA, & NCME, 2014). Tyto standardy, které jsou pravidelně aktualizovány⁴, jsou výsledkem spolupráce tří organizací – American Educational Research Association, American Psychological Association a National Council on Measurement in Education, a jejich první část se zaměřuje právě na vývoj testů. Bez povšimnutí by neměl zůstat ani soubor standardů organizace International Test Commission, z nichž jeden se věnuje procesu adaptace diagnostických nástrojů (ITC, 2005).

Nutné je však počítat s tím, že komplikace jsou přirozenou součástí procesu a mohou se objevit i přes sebepečlivější přípravu. S některými více či méně očekávatelnými překážkami jsme byli nuceni se vypořádávat i my v rámci projektu DIS. Vzhledem k tomu, že právě popis možných problémů v procesu vývoje či adaptace může být pro případné realizátory podobně orientovaných projektů nejužitečnější, pokusím se je v následujícím textu co možná nejméně tajit.

Před popisem konkrétních fází vývoje testu je vhodné upozornit na to, že budou pojímány z pohledu člena projektu DIS, který je specifický tím, že se během jeho realizace standardizovalo 12 diagnostických nástrojů téměř najednou. Což je fakt, který s sebou přináší zároveň výhody i nevýhody. Výrazným ulehčením je to, že v některých případech může být výstup vytvořený pro jeden test v mírné úpravě použitelný pro všechny ostatní (banálním příkladem je např. grafický design obálky). Kromě toho mohou některé fáze probíhat pro všechny testy najednou. Nejzajímavější je tato možnost

⁴ V současné době je k dispozici nová verze z července roku 2014. Do českého jazyka byla přeložena verze standardů z roku 1999, kterou je možné zakoupit v rámci vydavatelství Hogrefe Testcentrum, s.r.o. (AERA, APA, & NCME, 2001)

u procesu standardizace⁵, kdy je možné ověřovat několik testů současně na stejných respondentech, což může vést k zajímavým výzkumným zjištěním⁶. Prolínání fází vývoje bylo však paradoxně také největší nevýhodou projektu. Na stejném kroku ve vývoji (jako je např. analýza dat) u všech (či většiny) nástrojů se totiž často podíleli stejní lidé, kteří však mohli v daný moment pracovat pouze na jednom testu. Snažili jsme se proto vypracovat harmonogram tak, aby se tam, kde to nebylo účelné, stejné fáze pro různé nástroje pokud možno nepřekrývaly. To však v časově omezeném tříletém rámci projektu nebylo vždy možné.

Důvod a způsob vzniku testu

Charakter počátečních fází při vývoji nástroje se přímo odvíjí od toho, jaký má být jeho účel (zda má být součástí výzkumu, je odpovědí na požadavky praxe apod.) a jaký psychologický konstrukt má měřit. Oba tyto faktory pak mohou determinovat také způsob jeho vzniku. Základní dichotomií je v tomto kontextu adaptace na jedné straně a vývoj zcela nového testu na straně druhé. Toto primární dělení však dostatečně dobře nepokrývá celé spektrum nástrojů zařazených do projektu DIS – kromě zcela nových testů, které vznikaly přímo v rámci projektu a jejímiž autory se stali někteří z členů projektového týmu, byly do projektu zařazeny i nástroje, které se nacházely již v pokročilejší fázi svého vývoje. Jejich autory byli povětšinou odborníci přímo z praxe – např. psychologové či speciální pedagogové z pedagogicko-psychologických poraden, kteří vyhledali spolupráci nezbytnou ke standardizaci a vydání své metody. Další skupinou testů, které se způsobem práce v počátečních fázích podobají adaptacím, jsou revize starších českých nástrojů.

Každý výše uvedený typ testu má samozřejmě své výhody a nevýhody a přináší s sebou specifické problémy. Adaptace (případně také revize) obvykle poskytuje větší (mnohdy však falešný) pocit jistoty, protože nabízí již hotový rámec a posouvá výchozí bod vývoje testu blíže k cíli. Její největší a zásadní nevýhodou je však nutnost vypořádat se s licencemi (které obvykle, především u již praxí ověřených testů, drží komerční nakladatelství), což pro většinu výzkumných týmů představuje neřešitelný problém. Z pozice zástupce české státní instituce bývá bohužel někdy obtížné dosáhnout ve vyjednávání hned první fáze, a sice přimět zahraniční distribuční společnost ke komunikaci. Odměna za poskytnutí licence se pak obvykle řeší dvěma možnými způsoby (případně oběma variantami najednou) – odváděním částky za každý prodaný nástroj po jeho adaptaci, či jednorázovou platbou. První možnost je z důvodu malého trhu pro Českou republiku v podstatě vyloučena⁷, jednorázová částka za nákup licence bývá na druhou stranu natolik vysoká, že několikrát přesahuje zbytek rozpočtu celého

⁵ Termín standardizace je možné vymezovat různě. Např. Urbánek (2010) poukazuje na 3 odlišná pojetí tohoto pojmu. V rámci tohoto článku je standardizace chápána jako proces sběru dat pro tvorbu norem.

⁶ My jsme této příležitosti využili např. při vzájemné validizaci dvou inteligenčních testů.

⁷ V rámci projektu DIS nám bylo jedním překvapivě sdílným zástupcem zahraničního komerčního nakladatelství sděleno, že poskytování licencí k adaptaci nástrojů pro Českou republiku (s odváděním částky za každý prodaný nástroj) se nevyplatí čistě z praktických důvodů. Uhrazení právních služeb za vytvoření smlouvy totiž převyšuje potenciální zisk.

výzkumného projektu.⁸ Možným řešením je navázání spolupráce s některým ze dvou českých komerčních nakladatelství⁹, které mají pozici pro vyjednávání se zahraničím nepoměrně výhodnější. Tato spolupráce však může zásadním způsobem ovlivňovat průběh vývoje testu a je nutné ji zvážit především s ohledem na váš licenční vztah ke vzniklé adaptaci. V některých případech se však může stát jediným způsobem, jak do České republiky přinést kvalitní diagnostický nástroj.

Alternativou k adaptacím je vývoj zcela nového testu. Kromě problému s licenci v tomto případě odpadají i případné komplikace s kulturně závislými položkami (či celými subtesty), které je v některých případech nutné při adaptaci řešit. Problémy s kulturní vázaností však odpadají proto, že je třeba všechny položky teprve vytvořit. Záleží samozřejmě na komplexitě daného testu, ale pro vymezení obsahu jednotlivých subtestů a vytvoření konkrétních položek je obvykle nezbytná dlouhodobá spolupráce několika odborníků z dané oblasti. Vývoj zcela nového nástroje tak obvykle bývá procesem náročným odborně, časově i finančně.

I z těchto důvodů se nám v projektu DIS nejvíce osvědčil již zmíněný model spolupráce s českými odborníky. Příkladem může být diagnostický nástroj DISMAS¹⁰ (Traspe & Skalková, 2013), při jehož vývoji jsme spolupracovali s psychologem Pavlem Traspem a speciální pedagožkou Ivanou Skalkovou, kteří působí již několik let v pedagogicko-psychologické poradně Trutnov a mají dlouholetou praxi s dětmi s problémy v matematice. Prvotní nápad a vznik testu DISMAS byl reakcí přímo na potřeby poradenské praxe. Poté s pracovní verzí nástroje autoři několik let pracovali a experimentálně ověřovali a uzpůsobovali jeho obsah i formu tak, aby test co nejlépe vyhovoval svému účelu. Do spolupráce s autory jsme tak vstupovali s jistotou, že jde o nástroj, jehož koncepce přímo vychází z kontaktu s praxí, který nalezne své uplatnění a jehož forma i obsah byla dlouhodobě ověřována.

Je však zřejmé, že uvažování nad výše uvedenými skupinami a možnostmi může být do velké míry spíše teoretické, protože rozhodnutí pro konkrétní přístup k vývoji testu často vychází z potřeb a účelu vašeho výzkumného projektu.

Vznik první verze testu

Jak už bylo uvedeno, liší se charakter počátečních fází vývoje testu dle toho, zda se jedná o adaptaci, či se rozhodnete pro vytvoření zcela nového nástroje.

⁸ Pocit frustrace pak ještě umocňuje fakt, kdy se vašemu výzkumnému týmu velmi často podaří navázat úzkou spoluprací s autorem testu, který je adaptaci svého nástroje (a šíření povědomí o něm prostřednictvím nových výzkumných studií s důsledným citováním zdrojů) samozřejmě velmi nakloněn a vaše výzkumné snahy nadšeně podporuje. Problémem je, že práva ke svému testu prodal distribuční společnosti a do vyjednávání o licenci tedy nemůže obvykle nijak vstoupit.

⁹ Hogrefe Testcentrum, s.r.o. a Psychodiagnostika, s.r.o.

¹⁰ Nástroj DISMAS (Diagnostika struktury matematických schopností) mapuje rozvoj základních matematických schopností (se zaměřením na deficit v této oblasti) u předškoláků a dětí na prvním stupni základní školy. Skládá se z pěti subtestů – číselné řady, představy čísel, matematické pojmy, operační představy a automatizace.

V případě tvorby nového testu je po určení měřeného konstruktů naprosto zásadní přesně vymezení obsahovou doménu, která daný konstrukt pokrývá (např. jaké konkrétní znalosti a dovednosti zahrnuje konstrukt označený jako „matematické schopnosti“). V ideálním případě je tato obsahová doména podložena teorií. Pokud tomu tak je, stává se diagnostický nástroj její operacionalizací. Součástí těchto úvodních (nicméně naprosto zásadních) otázek je také výběr psychometrického modelu – bude test vznikat v rámci klasické testové teorie (CTT) nebo čím dál tím rozšířenější teorie odpovědi na položku (IRT)? Další důležité rozhodnutí (ovlivňující řadu následujících kroků) se týká formy testu – bude se jednat o klasičtější formu tužka-papír, nebo o počítačové testování? I v tomto případě je však nutné zopakovat, že odpovědi na tyto otázky mohou být do velké míry determinovány účelem testu. Po těchto základních vymezeních následuje tvorba designu testu – vytvoření konkrétních položek, sepsání pravidel administrace, určení způsobu skórování atd. Jak už bylo řečeno, jedná se o časově a odborně náročný proces a je vhodné do něj zapojit osoby, kteří se tématem, kterého se měření týká, dlouhodobě zabývají. Stejně tak je užitečné obrátit se na zkušené odborníky, pokud je testování zaměřeno na specifickou skupinu osob.

Pokud jde o adaptaci, je řada těchto otázek již vyřešena autory originálního nástroje. To by však nemělo znamenat, že bude celý rámec testu automaticky přejat bez zvážení smysluplnosti dílčích rozhodnutí, které původní autoři učinili, a jejich relevantnosti v novém prostředí. Namísto tvorby nových položek při vývoji nového testu je v rámci adaptace nezbytné zajistit jejich kvalitní překlad. S ohledem na kulturní podmíněnost testů by však byla chyba přistupovat k adaptaci pouze jako k převodu položek z jednoho jazyka do druhého, protože u některých z nich může překladem dojít k posunu od původního významu. Otázkou, se kterou jsme se potýkali i my v rámci projektu DIS, je pak v tomto kontextu užitečnost tzv. zpětného překladu¹¹. Tato metoda, která je řadou odborníků doporučována, může být vhodnou pomůckou, často však není samospasitelným řešením. Pokud totiž překladatel předem ví o tom, že bude jeho text podroben metodě zpětného překladu, může tento fakt (ať už nevědomky, či cíleně) ovlivňovat jeho volbu slov. Vzniklý text tak může vykazovat výborné výsledky z hlediska zpětného překladu, stylisticky však může působit kostrbatě. Navíc mají některá slova či slovní spojení napříč zeměmi a kulturami různé konotace. K vystižení podstaty položky (což by měl být primární cíl adaptace) tedy může být proto v některých případech vhodnější použít naprosto odlišná slova. Takový text pak samozřejmě nesplní podmínky zpětného překladu, i když může skvěle vystihovat význam originálu. Z těchto důvodů je dobré se zamyslet i nad jinými variantami metod k zajištění kvalitního překladu během adaptace. Jednou z nich může být zaslání překladu několika odborníkům, kteří nezávisle na sobě posoudí vhodnost volby slov a vytipují případné nepřesnosti či mnohovýznamnost u konkrétních položek. Pokud je to možné, je vhodné následně

¹¹ Zpětný překlad je jednou z metod, kterou se ověřuje správnost překladu v průběhu adaptace. V rámci této metody je text nejdříve přeložen odborníkem z originálního jazyka do jazyka vznikající adaptace. Následně je osloven druhý odborník (neobeznámený s nástrojem), který převede překlad zpět do původního jazyka. Tento vzniklý text se posléze porovnává s originálem. Případné výrazné rozdíly v textech mohou vypovídat o posunu ve významu při překladu a nutnosti úpravy položek.

uspořádat setkání těchto odborníků, v rámci kterého se dospěje k jednohlasnému závěru. V každém případě je užitečné, pokud je možné význam případných problémových položek diskutovat přímo s autorem originální verze nástroje či alespoň se zahraničním odborníkem, který má s testem dlouhodobější praktické zkušenosti.

Pilotáž a standardizace

Vzniklou pracovní verzi testu, ať už jde o nový nástroj či adaptaci, je nutné pilotně ověřit. Adaptace mohou svou povahou svádět k rezignaci a vypuštění pilotáže (a to např. i proto, že může být oproti tvorbě nového testu problematické – ať už kvůli licenční smlouvě či s ohledem na případné mezinárodní srovnávání výsledků – výrazně měnit či dokonce vypouštět položky, které vykazují po analýze dat z pilotáže nevyhovující parametry), nicméně nezbyvá než ji důrazně doporučit i v tomto případě. Pilotáž je vhodná, byť jen k ověření srozumitelnosti a především jednoznačnosti jednotlivých položek a instrukcí k administraci a vyhodnocení. I toto doporučení vychází z konkrétní zkušenosti. V rámci projektu DIS jsme se rozhodli pilotáž v případě adaptace komplexního nástroje IDS¹² (Grob et al., 2013) realizovat jen u dvou nových doplňkových subtestů¹³. Během standardizace se dle očekávání ukázalo, že téměř všechny subtesty (a s nimi spojené instrukce) fungují bez problémů. Jediné drobné komplikace se objevily u 4 subtestů mapujících socio-emoční kompetence a 1 subtestu zaměřeného na expresivní řeč. Na základě zpětných vazeb od odborníků z terénu se ukázalo, že by bylo v zájmu dosažení větší přesnosti a jednoznačnosti vhodné rozšířit u těchto subtestů překlad pokynů pro skórování odpovědí a doplnit ho konkrétními příklady. Vzhledem k tomu, že se změnou instrukcí (které byly doplněny na základě analýz velkého množství odpovědí ze standardizace) došlo také k mírnému posunu hodnocení některých reakcí, bylo nutné všechny položky těchto subtestů znovu projít a dle nového systému přebodovat. Kromě toho byl u 4 subtestů mapujících sociálně-emoční kompetence také v případě některých odpovědí dítěte nově kladen větší důraz na doplňující dotazy administrátora. Z tohoto důvodu nebylo možné některé původní odpovědi dětí ze standardizace adekvátně ohodnotit (protože během administrace nebylo provedeno nyní nezbytné doptání se dítěte). Po standardizaci byl proto proveden doplňkový sběr dat pro tyto 4 subtesty s cílem nahradit ztracená data a ověřit nově formulované instrukce k vyhodnocení. Pokud by byla provedena pilotáž (alespoň těchto pěti subtestů), byl by sice rozsah banky odpovědí, ze kterého bylo vycházeno při rozšiřování instrukcí ke skórování, podstatně menší, na druhou stranu by s největší pravděpodobností nebylo nutné provádět doplňkový sběr.¹⁴ Z těchto důvodů je vhodné

¹² Metoda The Intelligence and Development Scales (v české verzi IDS - Inteligenční a vývojová škála pro děti ve věku 5–10 let) je určena k diagnostice inteligence a obecné vývojové úrovně. V české verzi byla oproti originálu rozšířena o 2 verbální subtesty na celkových 21. Nástroj se zaměřuje na 6 oblastí vývoje – kognitivní předpoklady, motorické dovednosti, sociálně-emoční kompetence, matematické a jazykové schopnosti a výkonovou motivaci.

¹³ Toto rozhodnutí přitom vycházelo ze zvážení odborných i projektových konsekvencí. Někteří autoři adaptace měli již s přeloženou pracovní verzí nástroje praktické zkušenosti, které nesignalizovaly žádné komplikace, navíc bylo nutné brát ohled na časové a finanční možnosti projektu.

¹⁴ Aby byl dodržen slib z úvodu článku o snaze netajit vzniklé projektové komplikace, je nutné si položit otázku, zda může existovat nezanedbatelný rozdíl mezi výsledky v těchto čtyřech subtestech u dětí

nepodceňovat proces pilotáže i v případě adaptace a zaměřit se na získání detailních zpětných vazeb od účastníků sběru dat, na základě kterých pak může dojít k případným úpravám položek či instrukcí z manuálu. Kromě toho může pilotáž přinést důležité informace (např. spolehlivější odhad délky administrace testu) pro snadnější a přesnější plánování procesu standardizace. Během plánování harmonogramu je také vhodné zvážit možnost, že bude nutné provést pilotáž opakovaně. Především během vývoje zcela nového testu může být nezbytné provést několik na sebe navazujících cyklů pilotního ověřování.

Proces pilotáže a standardizace zahrnuje v zásadě stejné kroky a proto budou v této části článku popisovány najednou. Charakter jednotlivých kroků v rámci pilotáže a standardizace samozřejmě vychází z podoby nástroje, především z jeho komplexity a způsobu, jakým je administrován (tužka-papír oproti počítačové administraci např. přes internet). V případě testování online (nejčastěji prostřednictvím dotazníků či škál) postačuje obvykle jeden koordinátor, který spravuje elektronické rozhraní¹⁵ a komunikuje s respondenty. Následující text bude vycházet primárně z opačného pohledu – komplexního nástroje administrovaného formou tužka-papír, do jehož pilotáže či standardizace je nutné zapojit větší množství odborníků z celé ČR.

Právě výběr odborníků ke sběru dat je důležitým krokem v procesu pilotáže či standardizace. Při výběru a oslovování potenciálních sběračů je možné zvážit nastavení odborných kritérií, např. počet let strávených v poradenské praxi. I bez specifických požadavků však není shánění odborníků k zapojení do ověření testu snadnou záležitostí. I přesto, že měl projekt DIS v tomto ohledu situaci ulehčenou odbornou pozicí realizátora projektu (tedy instituce, která má vybudovanou kontaktní síť na odborníky ze škol a školských poradenských zařízení), bylo jednou z hlavních otázek procesu standardizace to, jak nejlépe motivovat potenciální sběrače dat a přesvědčit je k zapojení. Nakonec jsme stavěli hned na několika zdrojích motivace najednou. Jednalo se eticko-odborný apel („budete se podílet na vývoji kvalitního nástroje a zlepšení situace v rámci poradenské praxe“), odborné obohacení („naučíte se jako první v České republice pracovat s novým nástrojem, zdarma projdete školením v práci s ním a pokud to bude možné, nástroj vám zůstane jako odměna“) a samozřejmě finanční ohodnocení. Právě výši finanční odměny je nutné pečlivě zvážit. Administrace testu totiž nemusí být jedinou aktivitou, které se bude odborník v souvislosti s projektem věnovat. Bude test také vyhodnocovat? Jakým způsobem vám bude předávat výsledky? Bude sepsovat

z celého standardizačního souboru a dětí z doplňkového sběru. Jako příklad jedné z intervencí se nabízí vliv únavy z administrace celé baterie, který by zvýhodňoval děti z doplňkového sběru. K ověření této hypotézy by bylo třeba vytvořit kontrolní skupinu dětí, které by byla administrována celá baterie (včetně 4 subtestů s upravenými instrukcemi) a která by byla napárována dle důležitých charakteristik se skupinou dětí z doplňkového sběru. Výsledky těchto dvou skupin by pak byly srovnávány. I přesto, že předpokládáme, že je rozdíl mezi těmito skupinami zanedbatelný, výzkumné ověření této hypotézy by bylo vhodné uskutečnit. Bohužel ho však již nebylo možné realizovat s ohledem na časové a finanční omezení projektu.

¹⁵ Pro administraci online dotazníků je buď možné využít již existující webová rozhraní jako je např. Google Forms, LimeSurvey, SurveyMonkey, Survio apod., nebo v rámci projektu vyvinout nový systém přímo pro vaše potřeby.

zpětnou vazbu pro vás či rodiče zapojených žáků? Bude sám shánět respondenty do vzorku či zajišťovat individuální souhlasy? Některé tyto činnosti mohou být časově náročnější než samotná administrace testu a jejich neodměnění může vést k demotivaci a z ní vyplývajícím nepřesnostem či chybám v procesu sběru dat.

Dalším rozhodnutím, které je během plánování harmonogramu pilotáže a především standardizace nezbytné udělat, se týká počtu zapojených odborníků a stanovení množství testů, které mají administrovat. V rámci projektu DIS jsme vytvořili každému odborníkovi plán sběru, tedy minimální počet testů, které by měl za měsíc administrovat, aby byl splněn časový harmonogram. I přes toto opatření (rozfázování celkového vzorku do jednotlivých měsíců) se však ukázalo, že jsme se s některými odborníky domluvili na počtu, který z nejrůznějších důvodů (nepředvídatelné okolnosti jako je např. dlouhodobá nemoc, problém se sháněním respondentů do vzorku či prostě jen nereálně nastavený počet vyšetření) nebyli schopni splnit. To bohužel v případě některých testů vedlo k prodloužení doby sběru dat oproti původnímu plánu. Pokud je to možné a jste omezeni časovým rámcem projektu, nadhodnoťte raději počet odborníků zapojených do sběru dat. Samozřejmě je možné průběžně monitorovat narůstající vzorek a případně zapojit nové odborníky v průběhu standardizace, v takovém případě je však obvykle nutné počítat s tím, že bude třeba provést doplňkové školení.

Pečlivé zaškolení je, především u komplexnějších nástrojů, naprosto nezbytné. Jeho součástí by v ideálním případě mělo být také několik cvičných administrací, aby se odborníci s testy szili. I přes důkladné předání všech pokynů však vždy počítejte s určitým procentem chyb v administraci a nedodržení standardního postupu. Z tohoto důvodu je nezbytná kontrola odevzdaných záznamových listů. Ideální je provést ji v počátcích sběru dat po několika málo administracích, aby bylo možné případné nepřesnosti korigovat a nedocházelo k znehodnocování dat v dalších vyšetřeních. S tímto tématem se také pojí otázka autenticity získaných dat. V případě, že je během standardizačního procesu navázána spolupráce s externími odborníky, existuje možnost, že jsou jimi uváděné informace v záznamových listech nepravdivé? Nemusí se jednat o extrémní případ fiktivních probandů, ale o dodatečné zkreslení či úpravu některých dat, např. v případě zapomenutí otázky či subtestu během administrace. Ačkoliv jsme se s tímto problémem v projektu DIS nesešli (případně jsme ho neobjevili), je možné předpokládat, že šanci na neautenticitu dat může zvyšovat v předchozích odstavcích zmíněný velký časový tlak (nezvládnutí domluveného množství vyšetření) a pocit nedostatečného finančního ohodnocení.

Kromě nalezení a vyškolení odborníků pro sběr dat je nutné zaměřit svou pozornost na vyšetřované, tedy na vymezení vzorku, stanovení jeho velikosti a určení způsobu výběru respondentů vzhledem k měřenému konstruktovi a podchycení možných intervenujících proměnných. Realizační tým projektu DIS v tomto případě narazil se svými odbornými ideály na hranice proveditelnosti. Původní plán na zajištění co možná nejnáhodnějšího výběru (náhodný výběr dítěte ze seznamu všech žáků náhodně vybrané školy z celé ČR)

kombinovaný navíc se stratifikací vzorku dle dosaženého vzdělání rodičů a velikosti sídla (a samozřejmě obligátního věku a pohlaví) byl velmi rychle konfrontován s realitou. Komplikace se objevily hned v počátcích snah celý plán zkoordinovat. Problémem nebyly jen finanční možnosti projektu, velmi rychle se ukázalo, že není reálné pokrýt oblast celé republiky vyškolenými administrátory. Zredukovali jsme proto své ideály a modifikovali původní plán. Nový model, který se dařilo nějakou dobu realizovat, již nezahrnoval celou Českou republiku. Vymezili jsme sběrné oblasti v dojezdové vzdálenosti od místa působiště jednotlivých odborníků. V těchto několika kilometrových okruzích jsme pak náhodně vytipovávali školy k oslovení. Ani tento model však nakonec nebylo s ohledem na časový harmonogram a rozpočet projektu možné udržet. Kromě problémů s motivací ředitelů k účasti na projektu komplikovalo celý proces také dojíždění do vzdálenějších škol, které bylo pro odborníky často z časového hlediska nereálné. Z těchto důvodů jsme se nakonec museli vypořádat s nastoleným dilematem – buď stráví většinu svého projektového času odborníci na cestách do námi „náhodně“ vytipovaných škol a vyšetří za měsíc pětinu domluveného počtu žáků, nebo dodrží domluvený časový harmonogram, ale vyšetření bude probíhat na školách v místě jejich působení. S ohledem na časové omezení projektu jsme tedy znovu modifikovali model sběru dat a uchýlili se od metodologicky čistějšího náhodného výběru k pragmatičnosti příležitostného. Sběr tak sice probíhal ve všech krajích České republiky, ale odborníci obvykle vyšetřovali žáky ze škol, které se nacházely ve městech (či okolních vesnicích), ve kterých sběrači dat působili. Tento model byl výhodný nejen proto, že výrazně zredukoval čas strávený na cestách. Odborníci v minulosti již často s těmito školami spolupracovali, a proto bylo mnohem snazší přesvědčit vedení k zapojení do projektu.

V projektu DIS jsme si brzy uvědomili, že vývoj diagnostického nástroje je do velké míry sérií proseb a nabádání ke spolupráci. Po fázi motivačního působení na odborníky z terénu následuje (v minulém odstavci v krátkosti zmíněná) vlna žádostí směřující na ředitele škol a školek. I přesto, že jsme v některých případech volili třístupňový model motivačního nátlaku (dopis, e-mail, telefonát/y)¹⁶, podařilo se do spolupráce s námi zapojit přibližně každou pátou vytipovanou školu. Po získání souhlasu školy či školky je nutné motivovat ke spolupráci děti, respektive jejich zákonné zástupce. Celý proces je samozřejmě usnadněn, pokud projekt poskytuje možnost podpořit motivaci k zapojení odměnou. Kromě materiálního poděkování různého druhu se jako přirozená odměna nabízí poskytnutí zpětné vazby z vyšetření. V některých případech však může být vytvoření zpětné vazby časově náročnější než samotná administrace testu a do velké míry tak zvyšovat náklady spojené se standardizací. Kromě toho je problematické poskytovat zpětnou vazbu na základě výsledků testu, který teprve prochází procesem standardizace a nemá tudíž vytvořeny normy. Z těchto důvodů jsme rámci projektu DIS

¹⁶ V počátcích projektu jsme při oslovování škol s žádostí o spolupráci skutečně postupovali třífázově. Nejdříve jsme do školy zaslali doporučený dopis, po týdnu jsme e-mailovou cestou oslovili vedení školy a po několika dnech, v případě, že jsme nedostali odpověď, následoval telefonát (respektive telefonáty). V drtivé většině případů byl teprve telefonát podnětem, který vedl ke zvážení spolupráce (byť většinou se zamítavou odpovědí). U dopisu či e-mailu tedy často nešlo o odmítnutí žádosti, ale o její ignoraci.

spoléhali pouze na motivační sílu informačního dopisu a komunikační dovednosti třídního učitele, s jehož pomocí psycholog získával souhlasy k vyšetření. Metodologickou nepřijemností, která se v této souvislosti objevuje (a stejný problém se vyskytuje i na úrovni škol), je fakt, že poskytnutí souhlasu pouze pro některé žáky ze třídy devalvuje reprezentativnost vzorku tím, že neumožňuje náhodný výběr. Je pak samozřejmě možné předpokládat, že skutečně dochází ke zkreslení, protože k vyšetření ochotnější osoby mohou sdílet některé podobné charakteristiky. Celou tuto problematiku vystihl jeden z členů týmu glosou: „V demokracii prostě náhodné vzorky neexistují.“

Pokud není sběr dat realizován na počítači, následuje po něm (případně probíhá průběžně s ním) fáze přepisu výsledků ze záznamových listů do elektronické podoby. Ačkoliv se může kódování jevit jako banální proces, je problematický díky své poměrně velké náchylnosti k chybám. Pečlivý výběr kvalitního „přepisovače“ je důležitý vždy, u komplexních nástrojů, kde může převod jednoho záznamového listu do matice trvat i hodinu a standardizační vzorek dosahuje velikosti až 1500 respondentů, je však zásadní. Z těchto důvodů je nezbytné provést kontrolu prvních několika přepsaných záznamových listů. Na základě ní je pak možné korigovat další přepis, např. upřesnit kódování nejednoznačných položek, upozornit na nejčastější chyby apod. Pravidelnější kontroly Vám pak umožní, v případě, že je zapojených více přepisovačů, pro každého orientačně stanovit přesnost a rychlost přepisu (s případným vyvozením důsledků). Překvapilo nás, jak výrazně se mohou přepisovači v těchto charakteristikách lišit. Např. průměrná rychlost převodu jednoho záznamového listu nástroje IDS se mezi nejrychlejším a nejpomalejším přepisovačem lišila až o 40 minut (50 minut oproti 90 minutám)¹⁷. Chyby jsou však přirozenou součástí kódování a dochází k nim i přes sebepečlivější výběr přepisovače. Z těchto důvodů je vhodné podniknout takové kroky, které možnost chyb co nejvíce zredukuje. Jednou z takových možností může být nastavení datové matice tak, aby některý druh chyb přímo znemožňovala. Většina programů např. umožňuje nastavit charakteristiky hodnot, které je povoleno do matice zadávat¹⁸. Pokud je např. nutné přepsat odpovědi z posuzovací škály, jejichž hodnoty se pohybují v rozmezí 1–5, nastavíte podmínku pro vložení do buňky tak, aby jiné hodnoty nebylo možné vkládat. Tento druh chyby lze objevit i později po ukončení kódování pomocí jednoduchých deskriptivních statistik (např. tabulek četností či histogramů). Z logiky věci je však zřejmé, že tento postup neeliminuje všechny druhy chyb¹⁹. Z těchto důvodů může být vhodné přistoupit k náročnějším opatřením. Jedním z takových je např. metoda dvojího kódování²⁰, ve které jsou stejná data přepisována dvěma různými

¹⁷ Součástí záznamového listu IDS byly ještě další materiály a poznámky od odborníků z terénu. Kromě toho se přepisovaly doslovné odpovědi dětí. Z těchto důvodů se může doba strávená přepisem jednoho záznamového listu jevit jako nezvykle velká.

¹⁸ Například v programu Microsoft Excel tuto funkci naleznete v záložce „Data“ pod příkazem „Ověření dat“.

¹⁹ V angličtině bývají rozlišovány „out-of-range errors“ a „in-range errors“ (např. Barchard & Verenikina, 2013). Nastavením podmínek pro vkládání dat do matice či deskriptivními statistikami je přitom možné redukovat pouze první druh chyb.

²⁰ V angličtině bývá tato metoda označovaná jako „duplicate data entry“, „double entry verification“ či „two pass verification“.

osobami (v méně vhodném případě stejnou osobou v jiné době) do dvou odlišných matic, které se po ukončení přepisu elektronicky porovnají²¹. Jakékoli odlišnosti signalizují možnost chyby, kterou je nutné prověřit a případně opravit. Hlavní nevýhodou této metody, která umožňuje eliminovat velké procento chyb, je její náročnost na pracovní sílu, protože v zásadě zdvojnásobuje čas strávený převodem dat do elektronické podoby. Z tohoto důvodu jsme u některých testů v rámci projektu DIS používali mírně odlišnou metodu, která spočívala ve vizuální kontrole přepsaných dat (tedy porovnání matice a záznamových listů) jinou osobou než tou, která data přepisovala. Tato metoda vedla k objevení řady chyb a byla mnohem méně časově (a tím i finančně) náročná. V případě, že je však jediným kritériem k rozhodování přesnost a snaha eliminovat co nejvíce chyb, je samozřejmě nejvhodnější metoda dvojího kódování (např. Barchard & Pace, 2011; Barchard & Verenikina, 2013).²²

Po přepisu, důkladné kontrole a případné opravě (někdy se také mluví o čištění) dat přichází fáze jejich analýzy, která se liší dle zvoleného psychometrického modelu (CTT či IRT). Detailnější popis analýzy dat by však přesahoval rámec tohoto článku, a proto bude toto téma probráno jen stručně. Kromě základních deskriptivních statistik vzorku by měla práce s daty zahrnovat určitou formu položkové analýzy. V případě pilotáže se na základě jejích výsledků provádí revize testu, která pak vstupuje do standardizace (či další pilotáže). Kromě analýzy položek bývá ověřována dimenzionalita testu pomocí faktorové analýzy (ať už exploratorní, konfirmatorní či obou). Součástí jsou také analýzy spojené s reliabilitou a validitou vznikajícího nástroje a samozřejmě výpočet norem. Právě analýza dat byla jednou z fází, ve které projekt DIS nejvíce doplácel na skutečnost, že se v jeho průběhu standardizovalo 12 diagnostických nástrojů najednou. I přes fakt, že si projekt DIS mohl v této oblasti jen těžko přát vhodnější odborné zázemí, naráželi jsme během této etapy na problémy. Kromě toho, že bylo z důvodů neočekávaných komplikací nakonec nezbytné ukončit sběr dat pro několik testů v téměř stejnou chvíli a pracovat tedy paralelně na několika datových souborech najednou, ukázala se analýza časově náročnější, než jsme původně předpokládali. Situaci jsme se proto rozhodli flexibilně řešit pokusem o navázání externí spolupráce s dalšími psychometriky. I zde jsme však narazili na komplikace. Ty v tomto případě nespočívaly v absenci finančních prostředků projektu, ale v neexistenci samotných odborníků, kteří by disponovali potřebnými kompetencemi a zároveň dostatečným množstvím volného času. Buď je v České republice takových psychometriků nedostatek, nebo se nám je nepodařilo vypátrat. Částečně i toto podcenění počtu odborníků z této oblasti (ve spojení s velkou časovou tísní v závěru projektu) vedlo k tomu, že potenciál dat sebraných během standardizace testů nebyl využit stoprocentně. Získaná data sice splnila svou funkci při tvorbě norem a dalších nezbytných analýzách, ale v případě některých nástrojů se

²¹ I v tomto případě lze jednoduché srovnání provést přímo v programu Microsoft Excel, například pomocí funkce STEJNÉ.

²² Jako zajímavost lze ještě zmínit metodu čtení nahlas (reading aloud method), která je často používána v Japonsku (např. Kawado, Hinotsu, Matsuyama, Yamaguchi, Hashimoto, & Ohashi, 2003). K ní jsou zapotřebí dvě osoby – jedna čte nahlas již přepsaná data z datové matice (z počítače či z vytisknuté verze) a druhá porovnává čtené informace s daty na záznamovém listu.

podarilo vytvořit skutečně rozsáhlé matice vybízející k dalším zpracováním, které by mohly přinést zajímavé doplňující výsledky (ať už výzkumné, nebo přímo zlepšující kvalitu metody). Detailní explorace dat by však vyžadovala zapojení mnohonásobně většího počtu odborníků či podstatné navýšení doby realizace projektu. Doufejme, že nashromážděná data nezůstanou ležet ladem a bude s nimi dále pracováno i mimo projekt DIS.

Reliabilita a validita testu

Nezbytnou součástí vývoje diagnostického nástroje je úvaha nad způsobem ověřování jeho reliability a validity. Podcenění této problematiky může mít zásadní důsledky pro kvalitu vznikajícího testu. Ačkoliv je toto téma zmiňováno až nyní (téměř v samém závěru popisu vývoje diagnostického nástroje), je nezbytně nutné s ním spojené otázky vyřešit co nejdříve. Způsob výpočtu reliability a dokládání validity může totiž významným způsobem ovlivňovat jednotlivé fáze vývoje testu a to především sběr dat pro tvorbu norem.

Reliabilita je nejčastěji pojímána jako vnitřní konzistence. Je nicméně vhodné (pokud je to možné a vzhledem k charakteru daného nástroje smysluplné) zjišťovat i jiné její formy – test-retest, shodu posuzovatelů apod. Zajištění možnosti výpočtu těchto druhů reliability si však může vyžádat další finanční a časové prostředky a proto je nutné s nimi v harmonogramu vývoje testu počítat již od samého začátku.

Nezbytným procesem je vedle výpočtů reliability také dokládání důkazů o validitě vyvíjeného nástroje. Přitom je nutné pečlivě zvážit, jakým způsobem bude validita vznikajícího nástroje dokládána. Způsob dokládání validity, respektive pojetí validity jako takové se přitom v psychometrice postupně vyvíjí²³. Tradiční přístup (např. APA, 1966) rozlišuje validitu kriteriální²⁴ (dále dělenou na souběžnou a prediktivní), obsahovou a konstruktovou. V současné době se však pojetí různých typů validity opouští a na validitu se nahlíží spíše jako na jednotný, nečleněný koncept (např. Furr & Bacharach, 2014; Lissitz, 2009). Slosloví „typy validity“ nahrazuje „předkládání důkazů o validitě“. Stejně jako u reliability je nutné zvážit způsob dokládání validity s dostatečným předstihem, tedy ještě před sběrem dat pro tvorbu norem, protože může tento proces velmi významně ovlivňovat.

Je však nutné si uvědomit, že otázka validizace není spojena jen s fází sběru dat pro tvorbu norem, dokládání důkazů o validitě je průběžný (v zásadě nikdy nekončící) proces. Nejde o jednorázové zhodnocení testu, které má definitivní platnost (buď je, či není validní), ale spíše o posouzení toho, jak silné či slabé jsou předkládané argumenty (důkazy), které napovídají, do jaké míry jsou smysluplné a užitečné interpretace a intervence vyplývající ze získaných skóre testu. V ideálním případě by proto validizace testu neměla být ukončena jeho vydáním a jeho kvalita by měla být dále ověřována

²³ Tento vývoj je vysledovatelný např. v jednotlivých verzích již zmiňovaných Standardů pro pedagogické a psychologické testování.

²⁴ Někdy označována také jako empirická (Urbánek, Denglerová, & Širůček, 2011).

následnými validizačními studiemi. Pokud jde o nástroje z projektu DIS, aktuálně je např. chystána studie týkající se možnosti využití nástroje IDS u socio-kulturně znevýhodněných dětí. Rozšiřující studie, které je pak možné vydávat jako výzkumné články či doplňky manuálu, se však nemusí týkat jen validity. Vhodné by bylo i další ověřování reliability. U nástrojů z projektu DIS by se mohlo jednat například o formu test-rest, kterou z důvodu časového omezení projektu nebylo možné realizovat, o shodu posuzovatelů u posuzovacích škál testové baterie IDENA²⁵ (Hříbková, Nejedlý, Zhouf, & kol., 2013) apod.

Dokončení testu

Po ukončení fáze standardizace je vhodné využít zkušenosti odborníků, kteří s testem v terénu pracovali. V projektu DIS jsme v případě některých nástrojů k tomuto účelu vytvořili zpětnovazebný dotazník, který sloužil k zachycení připomínek či návrhů na úpravy různých oblastí standardizovaného testu. Kromě těchto dotazníků jsme plánovali uskutečnit setkání odborníků, kteří se podíleli na standardizaci nástroje IDS, jehož cílem mělo být nejen vzájemné předání zkušeností, ale také společná diskuse nad možnostmi doplnění či vylepšení testu²⁶. Změny provedené na základě těchto zpětných vazeb by se však samozřejmě již neměly týkat těch částí testu, které prošly standardizací. Po úpravách podnětového materiálu, instrukcí pro zkoumanou osobu či pravidel pro vyhodnocování by již nebylo možné použít získaná data a nástroj by musel projít standardizací znovu. Mělo by se tedy jednat o úpravy pouze těch částí testu či manuálu, které nemohou ovlivnit způsob odpovídání respondenta, či postup, jakým jsou jeho odpovědi vyhodnocovány.

V případě manuálu je ideální, pokud se na něm dařilo pracovat paralelně s vývojem testu a v tomto finálním stádiu se dokončují už jen ty jeho části, které nebylo možné zpracovat v průběhu předchozích fází vzniku nástroje. Jedná se zpravidla o popis průběhu standardizace a standardizačního vzorku, uvedení psychometrických charakteristik a výstupů analýz a sepsání částí zaměřených na vyhodnocení a interpretaci získaných skóre, které jsou na výsledcích zpracování dat závislé. Pro zajištění kvality nástroje je nezbytné, aby manuál obsahoval všechny potřebné informace o jeho charakteristikách a vývoji. V českém prostředí se podceňuje především uvádění psychometrických charakteristik a výsledků statistických analýz, tedy tzv. technická dokumentace diagnostického nástroje. Za zvážení stojí také možnost poskytnout jako součást testu matici s daty, na kterých byly prováděny veškeré analýzy. Každý, kdo bude chtít, tak získá možnost nahlédnout do nitra testu, ověřit správnost postupů a výpočtů a případně přijít s dalšími zajímavými výsledky.

Distribuce testu

Po posledních úpravách testu i manuálu nastává fáze vydání, v rámci které je nutné zvážit způsob a podmínky distribuce vzniklého nástroje. Otázky kdo, jak a za kolik je

²⁵ IDENA je souborem screeningových metod určeným k vyhledávání nadaných žáků na základních školách, který je tvořen 2 didaktickými testy (pro český jazyk a matematiku) a 7 posuzovacími škálami.

²⁶ Toto setkání se však již vzhledem k časové tísní v závěru projektu nepodařilo uskutečnit.

možné promýšlet i v případě, že byl test vytvářen z čistě výzkumných důvodů. V případě projektu DIS se odpovědi samy nabízely. Distributorem se u naprosté většiny testů stal řešitel projektu, tedy NÚV. Vzhledem k povaze financování a smyslu celého projektu jsou v rámci NÚV nástroje distribuovány za co nejnižší možné náklady. Kromě snahy o minimalizaci pořizovací ceny jsme se pokoušeli zredukovat i náklady spojené s užíváním nástroje. Z těchto důvodů je proto např. u testu DISMAS možné volně kopírovat spotřební materiál, tedy záznamové a vyhodnocovací listy.

S prodejem a šířením nástroje je spojena otázka zodpovědnosti distributora za správné užívání testu. Abychom zabránili používání nástrojů laiky, je nákup v rámci NÚV podmíněn doložením odpovídajícího (psychologického či speciálně-pedagogického) vzdělání. Diplom sám o sobě však samozřejmě nezaručuje absenci chyb v užívání diagnostické metody. Proto je u komplexnějších testů pro jejich nákup nutné absolvovat příslušný kurz, jehož cílem je předání zásad správné administrace, vyhodnocení a interpretace získaných výsledků.

Recenze testu

Smyslem projektu DIS však nebyl jen vývoj 12 diagnostických nástrojů. Se záměrem pomoci zavést v České republice tradici recenzního řízení jsme v rámci projektu využili překladu recenzního modelu EFPA (EFPA, 2009)²⁷ a nechali vzniklé diagnostické nástroje ohodnotit. K vyvážení projektové sebechvály je však nutné zmínit jednu věc. Základní podmínkou recenze je, aby byla vytvářena nezávislým odborníkem. Ačkoliv jsme přesvědčeni o tom, že je tak recenzenty nástrojů z projektu DIS možné označit, sluší se přiznat, že byli za své recenze placeni stejným projektem, který hodnocené diagnostického nástroje vytvářel. Věříme však, že tento fakt neměl vliv na nezávislost jejich posouzení.

Bylo by skvělé, pokud by se projekt DIS stal jedním z podnětů, které pomohou rozšířit povědomí o recenzním modelu, a recenzní řízení vzniklých diagnostických nástrojů se postupně stane běžnou součástí jejich vývoje. Nejde totiž jen o to, že poskytují důležité informace pro uživatele, jimž mohou sloužit jako základní vodítka pro zorientování se v nabídce testů. Jsou také důležitou zpětnou vazbou pro odborníky podílející se na vývoji metod, kteří mohou uvedené připomínky (např. na chybějící informace v manuálu, nedostatečné ověření validity či reliability nástroje apod.) využít při vývoji nové verze testu v budoucnu. Recenze by se tak mohly stát díky tlaku na kvalitu důležitým nástrojem při snaze o zlepšení psychologické diagnostiky v České republice.

Závěr

Tento článek stručně nastínil jednotlivá stádia vývoje diagnostického nástroje. Na otázky spojené s těmito fázemi odpovídal v kontextu zkušenosti s projektem DIS, který se během tří let zabýval vývojem 12 diagnostických metod. Právě v množství vyvíjených

²⁷ V tomto čísle časopisu Testforum vychází aktualizovaná verze recenzního modelu s novou formální úpravou.

testů spočívalo hlavní specifikum projektu DIS. Přesto věřím, že je výše uvedené zkušenosti alespoň částečně možné zobecnit i na podobné projekty menšího rozsahu. Důležitou částí těchto zkušeností, které se článek snažil zprostředkovat, byly komplikace, na které členové týmu projektu DIS v průběhu tvorby testů naráželi. Jednou z největších překážek, jejíž překonávání si vyžádalo nejvíce času a sil, byl napříč jednotlivými fázemi vývoje testů nedostatek lidských zdrojů. Vhodných osob potřebných k úspěšné realizaci daného stádia vývoje buď nebylo k dispozici dostatek (např. sběrači dat, psychometrici), nebo bylo nutné podněcovat jejich motivaci ke spolupráci (vedení škol, zákonní zástupci žáků apod.).

Díky snaze poukazovat na potenciální problémy při vývoji testů může celý článek vyznívat negativně či pesimisticky. Byla by však velká škoda, pokud by čtenáře odrazoval od podobně orientovaných výzkumných snah, protože jeho cílem byl přesný opak. Článek měl upozorněním na možné chyby napomoci k tomu, aby se pro zájemce z této oblasti stal vývoj testu zajímavým, podnětným a co nejméně stresujícím procesem, který bude ukončen vydáním kvalitního diagnostického nástroje.

Literatura

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2001). *Standardy pro pedagogické a psychologické testování*. Praha: Testcentrum.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Psychological Association. (1966). *Standards for educational and psychological tests and manuals*. Washington, DC: Author.
- Barchard, K. A., & Pace, L. A. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. *Computers in Human Behavior*, 27(5), 1834-1839.
- Barchard, K. A., & Verenikina, Y. (2013). Improving data accuracy: Selecting the best data checking technique. *Computers in Human Behavior*, 29(5), 1917-1922.
- European Federation of Psychologists' Associations (2009). Model recenze podle EFPA pro popis a hodnocení psychologických testů. Formulář recenze testu a poznámky pro recenzenty, Verze 3.41. *Testforum*, 0.
- Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics: an introduction*. Thousand Oaks: Sage Publications, Inc.
- Grob, A., Meyer, CH. S., Hagmann-von Arx, P., Krejčířová, D., Urbánek, T., Širůček, J., Jabůrek, M. (2013): *IDS - Inteligenční a vývojová škála pro děti ve věku 5–10 let*. Praha: Testcentrum.
- Hříbková, L., Nejedlý, P., Zhouf, J., & kol. (2013). *IDENA – posuzovací škály a didaktické testy k vyhledávání nadaných žáků*. Praha: Národní ústav pro vzdělávání.
- International Test Commission. (2005). *ITC Guidelines for Translating and Adapting Tests*. Retrieved September 22, 2014, from <http://www.intestcom.org/upload/sitefiles/40.pdf>
- Kawado, M., Hinotsu, S., Matsuyama, Y., Yamaguchi, T., Hashimoto, S., & Ohashi, Y. (2003). A comparison of error detection rates between the reading aloud method and the double data entry method. *Controlled clinical trials*, 24(5), 560-569.
- Lissitz, R. W. (Ed.). (2009). *The concept of validity: Revisions, new directions, and applications*. Charlotte: Information Age Publishing, Inc.
- Traspe, P., & Skalková, I. (2013). *DISMAS – Diagnostika struktury matematických schopností*. Praha: Národní ústav pro vzdělávání.
- Urbánek, T. (2010). Stav české psychologické diagnostiky a evropský model recenze testu. *Testforum*, 1(1).
- Urbánek, T., Denglerová, D., & Širůček, J. (2011). *Psychometrika: měření v psychologii*. Praha: Portál, s.r.o.