

Masarykova univerzita

Filozofická fakulta

Psychometrické aspekty psychologické diagnostiky:

Komentovaný soubor studií

Habilitační práce

OBSAH

1 ÚVOD	2
2 TESTOVÁNÍ OSOBNOSTNÍCH CHARAKTERISTIK	3
STUDIE 1 NEO osobnostní inventář 3: Základní psychometrické charakteristiky české verze NEO-PI-3	3
STUDIE 2 Big Five Inventory: Základní psychometrické charakteristiky české verze BFI-44 a BFI-10.....	5
STUDIE 3 Adaptivní administrace NEO PI-R: Výhody a omezení	6
STUDIE 4 Posouzení inkonzistence odpověďových vzorců v Eysenckově osobnostním inventáři.....	7
STUDIE 5 Jak přesné jsou národní stereotypy? Test různých metodologických přístupů	9
3 TESTOVÁNÍ KOGNITIVNÍCH SCHOPNOSTÍ	10
STUDIE 6 Testování prostorové představivosti: Konstrukce a zhodnocení nového nástroje	10
STUDIE 7 Inovativní testování prostorové představivosti: Interaktivní formát odpovídání a využití komplexního podnětového materiálu	11
STUDIE 8 Skryté aspekty v testování prostorové představivosti: Identifikace uplatňovaných stylů řešení položek.....	12
STUDIE 9 Experimentální porovnání tachistoskopických měření: Reakce na Wundtovu kritiku	14
4 ZÁVĚR	16
SUMMARY	17
PŘÍLOHY	18

1 ÚVOD

Předkládaný habilitační spis představuje komentovaný soubor publikovaných studií. Jednotlivým prvkem tohoto výběru je zaměření na psychometrické aspekty psychologické diagnostiky. Studie vznikaly v rámci různých výzkumných týmů, přičemž autorův příspěvek v projektech vždy spočíval v zajištění metodologických standardů v souladu s moderními trendy, ať již šlo o využití teorie odpovědi na položku, adaptivního testování, inovativních přístupů k měření specifických schopností či pokročilých postupů statistického zpracování dat při hodnocení kvality výzkumných nástrojů.

Soubor studií lze tematicky rozdělit do dvou základních oblastí. První oblast tvoří studie zaměřené na metody a postupy uplatňované v rámci zjišťování osobnostních charakteristik. Výzkumné úsilí bylo zaměřeno na adaptaci nejpoužívanějších osobnostních dotazníků založených na teorii Big Five, kde české odborné veřejnosti byla zpřístupněna standardizovaná verze komplexního osobnostního inventáře NEO-PI-3 a zejména pro vědeckou komunitu byla adaptována metoda BFI-44. Mimo adaptační studie byla teoretická poznatková báze v oblasti měření obohacena studiemi prozkoumávajícími možnosti adaptivního testování osobnostních charakteristik, potenciál matematického modelování odpověďových strategií, a také studií navrhuje nové statistické postupy pro posuzování přesnosti měření stereotypů v mezikulturních studiích.

Druhá oblast se zabývá problematikou testování kognitivních schopností. Zatímco ve výše zmíněné oblasti testování osobnostních charakteristik byl výzkumný zájem soustředěn na existující zahraniční metody, v případě testování kognitivních schopností byl vytvořen a psychometricky ověřen vícedimenzionální test prostorové představivosti, který pokrývá teoreticky identifikované faktory této schopnosti. Další zařazená studie oproti tradičně pojatému přístupu k testování prostorové představivosti představuje koncept počítačového testu, který aplikuje možnosti audiovizuální techniky v tvorbě inovativních položek, a hodnotí jejich přínos pro zvýšení ekologické validity testování. Třetí zařazená studie hlouběji prozkoumává mentální procesy zodpovědné za výkon v testech prostorové představivosti. Závěrečná studie propojuje psychometrické přístupy k měření kognitivních schopností od počátků vědecké psychologie až do současnosti a zaměřuje se na téma ekvivalence měření v kontextu možností a omezení replikace historických studií.

Předložené studie vznikly v rozmezí let 2011 až 2019 a byly z převážné části publikovány impaktovaných vědeckých časopisech. Jedna studie byla publikována v recenzovaném periodiku *Annales Psychologici*, čtyři studie v českém impaktovaném časopise *Československá psychologie*, a čtyři studie byly uplatněny v zahraničních impaktovaných časopisech (*American Journal of Psychology*, *Studia Psychologica*, *Cognitive processing*, a *European Journal of Personality*).

2 TESTOVÁNÍ OSOBNOSTNÍCH CHARAKTERISTIK

STUDIE 1 NEO osobnostní inventář 3: Základní psychometrické charakteristiky české verze NEO-PI-3

Odborná veřejnost se již dlouhou dobu intenzivně věnuje debatě o tom, na základě jakých vlastností lze nejužitečněji charakterizovat osobnost. Základní snahou výzkumníků při tom je identifikace základního setu dimenzí, který by byl co nejjednodušší, přičemž však neopomíjející důležité interindividální odlišnosti. Z hlediska statistického přístupu k psychologické realitě není překvapivé, že v rámci tohoto úsilí byly intenzivně využívány postupy vycházející z tradice faktorové analýzy. Právě tuto techniku lze totiž vnímat jako nástroj pro identifikaci latentní struktury základních faktorů, které stojí v pozadí komplexních projevů chování či prožívání.

Inventáře a dotazníky založené na teorii Big Five se postupně rozvíjí již 30 let. Pětifaktorový model osobnosti vychází z lexikálně-taxonomického přístupu, v rámci kterého bylo pomocí faktorové analýzy osobnostně relevantních přídavných jmen odvozena struktura základních pěti osobnostních dimenzí (neuroticismus, extraverte, otevřenost vůči zkušenosti, přívětivost a svědomitost). Podle názvů prvních tří vlastností byly metody nazvány NEO inventáře a pravidelně používány ve výzkumném i diagnostickém kontextu. Různé verze těchto inventářů byly přeloženy již do více než 37 jazyků a používány jsou ve více než 60 zemích. Velkou výhodou námi adaptovaného inventáře NEO-PI-3 je možnost interpretace osobnostní struktury nejen na úrovni základních pěti dimenzí, ale také v kontextu velmi podrobného profilu osobnostních vlastností na úrovni definovaných subškál. NEO-PI-3 se skládá z 240 položek, které jsou rozděleny do 5 základních škál, z nichž každá je následně rozčleněna na šest subškál, kde pro každou subškálu je vyhrazeno přesně osm položek. Položky jsou formulovány jak v pozitivním, tak i negativním směru, aby se předešlo systematickému zkreslení vlivem obecné tendence k souhlasu či nesouhlasu. Odpověďová škála inventáře je pětibodová s verbálním zakotvením *vůbec nevystihuje - spíše nevystihuje - neutrální - spíše vystihuje - úplně vystihuje*.

V rámci procesu české adaptace metody byla důkladná pozornost věnována všem aspektům standardizace psychodiagnostických metod. První verze českého překladu položek byla vytvořena na základě společné diskuse tří variant překladu od tří nezávislých překladatelů, která byla přeložena do angličtiny, a tento zpětný překlad byl zaslán autorům metody k posouzení možného posunu významu. Finální set položek byl administrován vzorku více než 1500 respondentů sestaveným na základě kvótního výběru. Získaná data sloužila k zevrubnému posouzení psychometrických charakteristik české verze metody.

Vnitřní konzistence byla zjišťována pomocí Cronbachova koeficientu alfa a lze konstatovat, že je obecně srovnatelná s originální (americkou) verzí metody, a také že došlo k pozitivnímu posunu

oproti předchozí verzi NEO-PI-R. Vynikající vnitřní konzistenci mají všechny základní škály instrumentu. Problematickou úroveň reliability ($\alpha < 0,70$) vykazovaly pouze čtyři subškály z celkového počtu třiceti, což však zjevně není způsobeno kvalitou českého překladu, neboť podobně nízké hodnoty se objevují také u americké a německé verze metody.

Nejdůležitějším analytickým nástrojem v kontextu Big Five je faktorová analýza, přičemž se upřednostňuje explorační faktorová (či komponentová) analýza metodou hlavních komponent s rotací Varimax. Vzhledem k vysokému počtu položek se faktorová analýza prováděla na úrovni subškál a bylo sledováno, zda se subškály adekvátně sdružují do škál odpovídajících základním osobnostním dimenzím. Výsledky faktorové analýzy lze považovat za zcela uspokojivé, neboť žádná subškála nevykazovala sekundární faktorový náboj vyšší než náboj zjištěný na primárním faktoru. Pozornost byla věnována také kongruenci české verze NEO-PI-3 s verzí jak americkou NEO-PI-3, tak i s českou předchozí verzí metody NEO-PI-R. Zvolili jsme v literatuře popsanou proceduru, kdy jsme danou faktorovou strukturu (v našem případě např. strukturu subškál české verze NEO-PI-3) rotovali do struktury, která je již předem určena a slouží jako „cílová matice“, které jsme se Procrustovou rotací snažili přiblížit. Faktorové náboje rotovaného řešení pak byly srovnány s normativními hodnotami pomocí koeficientů kongruence. Výsledky tohoto postupu jednoznačně hovoří o vysoké úrovni kongruence s oběma referenčními verzemi metody. V případě NEO-PI-R se koeficienty kongruence pro jednotlivé subškály pohybovaly v rozmezí 0,92 až 1,0, pro jednotlivé škály pak v rozmezí 0,98 až 0,99. Celkový koeficient kongruence měl hodnotu 0,98. Podobně vysoké hodnoty byly nalezeny také v případě porovnání s americkou verzí NEO-PI-3 (pro subškály 0,92 a 0,99, pro škály 0,97 až 0,98, celková kongruence 0,98).

V rámci předložené studie nebyla zjišťována kriteriální validita metody, nicméně vzhledem k výrazné kongruenci se starší verzí NEO-PI-R se lze při jejích odhadech opřít o předchozí studie, které obecně hovoří o velmi dobré úrovni validity. Pro klinicko-diagnostické účely byly pro metodu NEO-PI-3 vytvořeny i normy ve formě percentilů a T skóre, a to odděleně pro různé věkové skupiny a pohlaví, které byly publikovány v manuálu¹ metody.

Bibliografická citace: Hřebíčková, M., & Jelínek, M. (2019). NEO osobnostní inventář 3: Základní psychometrické charakteristiky české verze NEO-PI-3. *Československá psychologie*, 63(2), 160-176.

¹ Hřebíčková, M. & Jelínek, M. (2016). Česká adaptace inventáře NEO-PI-3 autorů: P. T. Costa a R. R. McCrae. Praha: Hogrefe - Testcentrum.

STUDIE 2 Big Five Inventory: Základní psychometrické charakteristiky české verze BFI-44 a BFI-10

Dalším nástrojem z rodiny Big Five, který byl adaptován do českého prostředí, je BFI-44, respektive jeho zkrácená verze BFI-10. Číslovka v názvu metod odkazuje na celkový počet položek, a je tedy zřejmé, že se jedná o poměrně krátké nástroje. V případě BFI-10 jde o jeden z nejkratších dotazníků sloužících ke komplexnímu postihnutí osobnostních charakteristik. Na rozdíl od NEO-PI-3 jde o volně dostupné nástroje, což je společně s jejich kvalitními psychometrickými charakteristikami předurčilo k široké oblibě ve výzkumném prostředí. Další výhodou BFI dotazníků je záměrně jednoduchá formulace položek, která umožňuje uplatnění v rámci širokého věkového rozpětí.

Big Five Inventory byl v různých překladech výzkumně používán již dříve, a tak hlavním cílem této studie bylo poskytnout odborné veřejnosti jednotný překlad položek, ke kterému by byly dostupné ověřené psychometrické charakteristiky. Standardizace nástroje byla provedena na rozsáhlém souboru respondentů (celkem více než 6000 osob) ve věkovém rozpětí od 11 do 84 let. Celkový soubor respondentů se skládal z šesti dílčích podsouborů, které byly získány v rámci různých výzkumných projektů. Abychom ověřili konzistenci získaných výsledků a mohli porovnat výsledky pro dospělou a adolescentní populaci, analýzy byly provedeny odděleně v jednotlivých podsouborech.

V dílčích analýzách jsme se zaměřili na reliabilitu ve smyslu vnitřní konzistence nástrojů a faktorovou strukturu na úrovni položek. Vzhledem k tomu, že jsme měli k dispozici data z opakovaného měření, bylo možné posoudit také stabilitu nástroje. Z důvodu extrémně nízkého počtu položek připadajících na jednotlivé dimenze u zkrácené verze BFI-10 byla zjišťována také schopnost škál reprezentovat škálové skóry z nezkrácené verze BFI-44 na základě výpočtu množství sdíleného rozptylu. Souhrnně lze konstatovat, že nástroje prokázaly uspokojivé hodnoty vnitřní konzistence (a to i u vzorku adolescentních respondentů) i test-retest reliability. Faktorová struktura položek BFI-44 v podstatě odpovídá teoretickému očekávání o rozložení do pěti základních dimenzí, ačkoli bylo patrné, že u respondentů v rané adolescenci nástroj vykazoval větší počet problematických položek. Hodnoty míry rekonstrukce svědčící o kvalitě zkrácené verze nástroje jsou poměrně dobré, neboť ačkoli BFI-10 obsahuje pouze 25 % položek plné verze BFI-44, přesto předpovídá 57 % rozptylu plné verze. Závěrem lze konstatovat, že české adaptace BFI-44 a BFI-10 zejména při sebeposouzení dospělými respondenty mají akceptovatelné psychometrické vlastnosti, reprezentované přiměřenou vnitřní konzistencí škál, odpovídající retestovou reliabilitou, a adekvátní faktorovou strukturou. Nicméně je třeba poznamenat, že o krátké verzi BFI-10 je vhodné uvažovat pouze při kritickém nedostatku času, a pokud má výzkumník více prostoru při administraci testové baterie, je vždy vhodnější použít úplnou verzi se 44 položkami.

Bibliografická citace: Hřebíčková, M., Jelínek, M., Blatný, M., Brom, C., Burešová, I., Graf, S., ... Zábrodská, K. (2016). Big Five Inventory: Základní psychometrické charakteristiky české verze BFI-44 a BFI-10. *Československá psychologie*, 60(6), 567–583.

STUDIE 3 Adaptivní administrace NEO PI-R: Výhody a omezení

Praxe vytváření zkrácených verzí psychodiagnostických nástrojů svědčí o tom, že ve výzkumné i klinické praxi je velký zájem o zvýšení efektivity testování ve smyslu snížení zátěže klienta či respondenta. Tato praxe je obvykle založena na identifikaci obecně psychometricky kvalitních položek, a to zejména ve smyslu rozlišovací účinnosti. V rámci tohoto přístupu však nejsou dostatečně zohledněny individuální specifika testovaných osob. Odlišný přístup je uplatněn v rámci adaptivního testování, které je však nejčastěji využíváno v oblasti výkonových testů s dichotomně skórovanými položkami. Základem adaptivního testování na principu teorie odpovědi na položku je odhad informačního přínosu konkrétní položky pro jedince s konkrétní úrovní měřeného rysu. V praxi se postupuje iterativně tak, že v každém kroku adaptivní administrace je z vektoru dosavadních odpovědí odhadnuta úroveň měřeného rysu, na jejímž základě je vybrána taková následující položka, která o jedinci sdělí maximum informace. Postupně je s přibývajícím množstvím odpovědí na administrované položky upřesňován odhad rysu, dokud nedojde k dosažení požadované přesnosti měření. Efektivita takového postupu byla opakovaně prokázána v množství relevantních studií.

V principu nic nebrání aplikaci adaptivního přístupu i na metody využívající položky polytomního formátu, které se velmi často objevují v oblasti diagnostiky osobnosti. Pro klasické škály Likertova typu, které se používají v osobnostních dotaznících, se zdá být nejvhodnější model odstupňovaných odpovědí (GRM). Výhoda modelu spočívá v tom, že není kladeno žádné omezení na počet odpověďových kategorií v rámci položek, jednotlivé položky určitého testu se dokonce mohou lišit v počtu použitých kategorií. Model GRM je dobře konceptuálně navržený, srozumitelný a v zásadě vhodný pro většinu nástrojů v oblasti psychologické diagnostiky. Z těchto důvodů jsme model GRM zapracovali do software CATo². Pro zjišťování efektivity a ověření ekvivalence adaptivní a klasické administrace NEO-PI-R byl použit postup simulace z reálných dat, která místo postupného předkládání položek skutečné osobě získává odpovědi z předem sestavené databáze obsahující odpovědi osob z klasické administrace. Vstupní data pocházela ze standardizačního projektu metody a obsahovala více než 2000 vektorů odpovědí. Adaptivní administrace pro simulované průchody testem byla nastavena tak, že výpočet úvodního odhadu rysu byl proveden na základě odpovědí na tři

² Software CATo (Computerized Adaptive Testing optimized) je nástroj pro adaptivní administraci testů, který byl vyvinut týmem výzkumníků na půdě Psychologického ústavu Akademie věd ČR (Martin Jelínek, Petr Květon, Denisa Denglerová). V software jsou implementovány modely pro dichotomní i polytomní položky a je schopen také administrace inovativních položek s audiovizuálním obsahem.

náhodně vybrané položky, každá následující položka byla vybrána na základě principu maximálního informačního přínosu, a administrace byla ukončena při dosažení chyby odhadu nižší než 0,45, což odpovídá obecně akceptované úrovni reliability 0,80. Pro drtivou většinu osob byl potřebný počet položek velmi nízký ve srovnání s plnou délkou jednotlivých škál. Mediánový počet se pohyboval v rozmezí 8 (neuroticismus) až 11 (otevřenost a přívětivost) položek. Pouze ve velmi malém množství případů byla daná škála administrována v plné délce (48 položek), když v průběhu administrace nedošlo ke splnění podmínky dosažení dostatečné přesnosti měření. Nejrelevantnějším kritériem pro posouzení, zda adaptivní administrace dostatečně rekonstruuje výsledky získané na základě neredukované škály, je míra vztahu mezi odhady rysu za použití aparátu IRT. V této simulační studii bylo ověřeno, že míra rekonstrukce dosahuje vynikající hodnoty 87 % při zkrácení testu přibližně o dvě třetiny délky. I přes slibné ukazatele efektivity tato studie poukázovala i na limity adaptivního testování, a to zejména problém nadužívání psychometricky nejkvalitnějších položek.

Bibliografická citace: Jelínek, M., Květon, P., & Vobořil, D. (2011). Adaptivní administrace NEO PI-R: Výhody a omezení. *Československá psychologie*, 55(1), 69–81.

STUDIE 4 Posouzení inkonzistence odpověďových vzorců v Eysenckově osobnostním inventáři

V diagnostické praxi je běžné, že se odborník při stanovování diagnózy opírá o výsledky standardizovaných dotazníků. Nicméně odborníci jsou si samozřejmě vědomi limitů a omezení pro interpretaci výsledných dotazníkových skóre. V oblasti diagnostiky osobnosti se uživatelé metod setkávají zejména s problémem odhalení náhodného či nedbalého odpovídání, který může být více častý u specifických skupin respondentů, například dětí či adolescentů. I proto je obvyklé, že manuály metod odkazují na důležitost vizuální kontroly odpověďových archů, případně jsou zařazeny různé formy validizačních škál. Další zdroj zkreslení vyplývá ze samotné povahy měřících nástrojů, které jsou většinou postaveny na nomotetické povaze sledované charakteristiky odhalené za pomocí pokročilých statistických postupů (faktorová analýza). Současně je ale třeba poznamenat, že při využití takového přístupu nelze očekávat, že chování každého jedince (jeho individuální vzorec odpovědí) bude bezesbytku v souladu s teoretickým očekáváním (v cizojazyčné literatuře je používán termín *lack of traitedness*, tj. nedostatek rysovosti). Teorie odpovědi na položku poskytuje elegantní statistický aparát, pomocí kterého lze indikovat nekonzistentní odpověďové vzorce v rámci jednotlivých dimenzí osobnosti.

V této studii byly aplikovány dva výpočetně odlišné indexy (Z_3 a F_2) na datovou matici s odpověďmi na položky dotazníku EPI (se škálami neuroticismus a extraverte) od více než 400 adolescentů, přičemž cílem bylo porovnat výsledky obou indexů, identifikovat osoby s vyšší inkonzistencí v odpovídání a analyzovat možné příčiny zjištěné inkonzistence. Z výsledků studie vyplynulo, že ani jeden z indexů není ovlivněn celkovou úrovní rysů a podávají v zásadě shodnou informaci. Pro identifikaci inkonzistentních vzorců odpovědí byl využit Z_3 index, neboť jeho hodnoty lze interpretovat v rámci standardního normálního rozložení, kdy nízké hodnoty (menší než -2) indikují závažnou odchylku od normativně očekávaného vzorce. V případě zjištěné odchylky se pak těžiště úvah přesouvá ke zjištění příčin inkonzistentního odpovídání. V principu se může jednat buď o náhodné odpovídání či o individuální odchylky v rysovosti osoby. V kontextu těchto úvah lze aplikovat dva postupy. První z nich spočívá v porovnání s validizačními škálami. Ačkoli dotazník EPI obsahuje tzv. lži-škálu, nelze v tomto případě škálu využít, neboť není zaměřena na inkonzistenci v odpovídání, ale pouze na sociální desirabilitu. Proto byl ve studii použit druhý postup spočívající v porovnání hodnot indexu napříč různými škálami dotazníku. Vzhledem k tomu, že položky obou škál jsou v dotazníku smíchány dohromady, lze předpokládat, že pokud je zjištěna inkonzistence u stejného respondenta v obou škálách, je pravděpodobným zdrojem zkreslení náhodné odpovídání. Pokud však dojde ke zkreslení pouze v jedné ze škál, je považována za pravděpodobnější interpretace z hlediska nedostatku rysovosti.

Způsob interpretace představených indexů (patřících do skupiny tzv. person-fit indexů) závisí na kontextu testové situace. Vědec, který pracuje s velkým množstvím dat, je může použít k identifikaci obtížně interpretovatelných (nekonzistentních) případů, aniž by se hlouběji zabýval příčinami. Jiná situace nastává v klinickém prostředí. Zkušený psychologický odborník ví, že stejné skóre dotazníku neznámá vždy stejnou kvalitu pro dva různé klienty. Pomocí person-fit indexu se může ujistit, že skóre může interpretovat v souladu s teoretickým pozadím metody či zda je nutné k interpretaci přistupovat obezřetně.

Bibliografická citace: Jelínek, M., Květon, P., & Vobořil, D. (2010). Assessment of response pattern aberrancy in Eysenck Personality Inventory. *Annales Psychologici*, 58(14), 37–46.

STUDIE 5 Jak přesné jsou národní stereotypy? Test různých metodologických přístupů

Problematika národních stereotypů je důležitým aktuálním tématem na pomezí sociální psychologie a psychologie osobnosti. Zajímavou a stěžejní otázkou je, nakolik lze národní stereotypy považovat za přesné, tj. jak moc spolu souvisí národní autostereotyp (posouzení typického představitele vlastního národa) a sebeposouzení skutečných osob v rámci konkrétního národa. Prezentovaná studie využívala teoretické pozadí Big Five a byla provedena v pěti státech střední Evropy a pěti státech baltského regionu. Více než 17 tisíc respondentů poskytlo odpovědi na dotazník NEO-PI-R nebo NCS (National Character Survey - nástroj vytvořený specificky za účelem popisu typického příslušníka určité kultury v kontextu Big Five).

Obecným cílem studie bylo zjistit, zda odlišné metodologické přístupy k designu studie a analýze dat ovlivňují zjištěnou úroveň shody mezi posouzením národního stereotypu a sebeposouzením. Konkrétně šlo o zjištění vlivu použité metody sebeposouzení (NEO-PI-R vs. NCS), dále o rozdíly způsobené využitím rozdílných norem používaných v předchozích studiích (ruských, mezinárodních, nebo amerických) pro konverzi hrubých skóreů na standardizované T-skóre, a rozdíly způsobené využitím odlišného statistického aparátu (vnitrotřídní korelace, Pearsonovy korelace, a Spearmanovy pořadové korelace). Výsledky studie dokazují, že interpretace studií národních stereotypů nemohou být prováděny nezávisle na použitých metodologických postupech. Z hlediska statistických postupů se vede dlouhodobá debata o vhodnosti použití Pearsonova koeficientu korelace či koeficientu vnitrotřídní korelace, neboť ačkoli oba koeficienty bývají používány ke stejnému účelu, neposkytují ekvivalentní údaje a navíc z konceptuálního hlediska trpí oba koeficienty při použití v této oblasti výraznými nedostatky. Navíc doposud nebyla spolehlivě vyjasněna procedura určení statistické významnosti těchto koeficientů. Ve studii byl navržen alternativní postup určení shody sebeposouzení a národního stereotypu. Zatímco Pearsonův koeficient a koeficient vnitrotřídní korelace patří do skupiny parametrických postupů, prezentovaný nový postup využívá neparametrickou pořadovou korelaci aplikovanou na pozici zemí dle sebeposouzení i národního stereotypu. Problém s nejasností určení statistické významnosti výsledků v předchozích studiích využívajících klasické postupy byl překonán aplikací techniky bootstrappingu, která spočívá ve vytváření mnohonásobných opakovaných náhodných vzorků dat. Pro další směřování výzkumu v oblasti přesnosti národních stereotypů studie na závěr poskytuje praktická doporučení založená na zjištěných skutečnostech.

Bibliografická citace: Hřebíčková, M., Möttus, R., Graf, S., Jelínek, M., & Realo, A. (2018). How accurate are national stereotypes? A test of different methodological approaches. *European Journal of Personality*, 32(2), 87–99.

3 TESTOVÁNÍ KOGNITIVNÍCH SCHOPNOSTÍ

STUDIE 6 Testování prostorové představivosti: Konstrukce a zhodnocení nového nástroje

Koncept prostorové představivosti je považován za důležitou součást kognitivních předpokladů k úspěšnému fungování v mnoha oblastech lidské činnosti, nicméně je poměrně obtížné ho definovat a následně i operacionalizovat. Po dlouhou dobu byla prostorová představivost považována za součást konstruktů obecné inteligence, ale od počátku 20. století je spíše vyčleňována jako samostatná schopnost. Po osamostatnění prostorové představivosti jako svébytné kognitivní schopnosti se vědecká debata zaměřila na vnitřní strukturu tohoto konstruktů a polemika se vedla o množství a povaze potenciálních komponent. Námi vytvořený nástroj, představený v této studii, se opírá o Maierovu kategorizaci, ve které jsou vyčerpávajícím způsobem zkombinovány různé teoretické přístupy do podoby pětifaktorového modelu. Tento model zahrnuje prostorovou percepci, prostorovou vizualizaci, mentální rotaci, prostorové vztahy a prostorovou orientaci, které však nepředstavují zcela nezávislé komponenty, ale vykazují dle autora vysokou míru vzájemné provázanosti. Pro jednotlivé komponenty bylo na základě rešerše existujících nástrojů navrženo dohromady pět prototypických položek, které byly podrobeny obsahové validizaci deseti experty. Na základě jejich hodnocení bylo ověřeno, že prototypy skutečně odrážejí zamýšlené dimenze prostorové představivosti, nicméně u dvou prototypů byl zjištěn přesah i do sekundární dimenze. Tento fakt byl zohledněn v následujících analýzách. Od každého prototypu bylo odvozeno dalších šest položek, které mezi sebou byly rozlišeny stupněm obtížnosti. Celkově tak byla získána položková banka čítající 35 položek.

Psychometrické charakteristiky položek byly ověřeny na vzorku 267 respondentů, kterým byl test administrován v online prostředí. Získaná data byla podrobena sérii konfirmačních analýz. Konkrétně byl postulován model s pěti korelovanými faktory, hierarchický model s obecným faktorem prostorové představivosti a pěti komponentami, a konečně jednodimenzionální model. Na základě posouzení indexů shody s daty a hodnot regresních koeficientů byl shledán jako nejvíce vystihující hierarchický model. Následná položková analýza tedy byla provedena jak na úrovni celkového skóru, tak na úrovni jednotlivých subškál. Bylo zjištěno, že položková banka neobsahuje žádnou špatně fungující položku, nicméně z hlediska rozlišovací účinnosti bylo možno některé položky eliminovat a test zkrátit při zachování dostatečné reliability. Pro zachování obsahové vyváženosti testu jsme se rozhodli z každé dimenze ponechat stejný počet pěti položek. Finální test byl administrován vzorku 124 studentů v rámci studie využívající kompletní ekvivalenční design umožňující zachytit vliv média administrace (papír vs. online), test-retest stabilitu a vzhledem k zařazení standardizovaného subtestu Prostorová představivost z Testu struktury inteligence IST-70 také konvergentní validitu. Výsledky ukázaly dobrou úroveň reliability (jak ve smyslu stability, tak vnitřní konzistence), nezávislost na

formě administrace, a uspokojivou validitu testu. Závěrem lze konstatovat, že test SRT (Spatial Reasoning Test) je vhodným a psychometricky kvalitním nástrojem k zachycení schopnosti porozumět prostorovým vztahům a mentálně manipulovat s objekty v prostoru.

Bibliografická citace: Květon, P., Jelínek, M., & Vobořil, D. (2014). Testing of spatial ability: Construction and evaluation of a new instrument. *Studia Psychologica*, 56(3), 233–251.

STUDIE 7 Inovativní testování prostorové představivosti: Interaktivní formát odpovídání a využití komplexního podnětového materiálu

Test SRT představený v předchozí studii lze označit za klasický test v tom smyslu, že se skládá z jednoduchých položek tradičního formátu, kdy každá úloha obsahuje statické zadání a v odpověďové části nabízí několik distraktorů a jednu správnou odpověď. Pro administraci testu tedy není nutné využívat pokročilých technologií, ale postačí formát papír/tužka. V této studii jsme prozkoumali možnosti inovativního testování, které je založeno na možnostech, které nabízí počítačová administrace testu. Obecně lze říci, že inovativní princip lze uplatnit jak v zadání položky (nejčastěji formou audiovizuálních prvků), ale i ve formátu odpovídání. Navrhnuty byly dva koncepty inovativních testů, které byly různé jak formátem inovativního položkového principu, tak i zaměřením na odlišné aspekty prostorové představivosti.

Palčivým problémem při tvorbě klasických výkonových testů je mj. tvorba adekvátních distraktorů, jejichž kvalita pochopitelně zásadním způsobem ovlivňuje výsledné parametry každé položky. Bývá velmi obtížné sestavit sadu distraktorů srovnatelných z hlediska jejich atraktivity a současně udržet tuto atraktivitu natolik vysokou, aby distraktory nebylo možné vyloučit i bez požadované úrovně měřené schopnosti. Formát prvního testového konceptu (ve studii jde o Hook test) je inovativní ve způsobu odpovídání, kdy zkoumaná osoba má za úkol interaktivně upravit zadání tak, aby se co nejvíce přiblížilo požadovanému cílovému stavu, který by byl správnou odpovědí v klasickém testu. Měřena je zde numericky vyjádřená odchylka od tohoto cílového stavu, což s sebou přináší další výhodu z hlediska teorie měření, neboť výkon je kvantifikován na vyšší, poměrové úrovni měření. Vzhledem k předpokládané spojitě povaze sledované schopnosti lze tuto míru považovat za vhodnější. Validita konceptu byla ověřena korelací s klasicky administrovaným testem s pevným odpověďovým formátem. Nezanedbatelnou výhodou takto koncipovaného testu je efektivita při tvorbě testových položek, neboť prakticky odpadá problém s navrhováním distraktorů a ověřováním jejich kvality.

Druhý navržený koncept (CityRide test) v daleko větší míře využívá inovativních prvků a je zaměřen zejména na posílení ekologické validity procesu testování, což je v oblasti prostorové představivosti hojně diskutovaná problematika. Jak již název napovídá, CityRide test pokrývá komplexní schopnost prostorové orientace. Jeho základem jsou videozáznamy průjezdu reálným městským prostředím z pohledu řidiče osobního automobilu. V konceptu se prozkoumávaly dvě varianty odpovídání, kdy první z nich v zásadě odpovídala klasické testové formě (výběr správného profilu projeté trasy z pěti možností), zatímco druhý byl ve své podstatě interaktivní, neboť úkolem respondenta bylo graficky určit azimut vzájemné polohy startu a cíle. Z hlediska psychometrických kvalit se jednoznačně ukázal jako více slibný plně interaktivní formát testu. Domníváme se, že tento položkový koncept vyžaduje zapojení analogových strategií řešení položek (které jsou skutečným podkladem prostorové představivosti, viz Studie 8), a v podstatě vylučuje vypomáhání si při řešení verbálně-analytickými strategiemi, které poměrně výrazně interferují s výkonem v klasických testech prostorové představivosti. Vzhledem ke komplexní povaze úloh v CityRide testu, jejichž řešení zjevně vyžaduje zapojení i dalších kognitivních schopností nad rámec prostorové představivosti, doplnili jsme design studie také o měření základních složek pracovní paměti, tj. verbální pracovní paměti a vizuálně-prostorové pracovní paměti. Výkon v CityRide testu pozitivně koreloval s vizuálně-prostorovou pracovní pamětí, zatímco vztah s verbální pracovní pamětí nebyl nalezen, což lze interpretovat ve smyslu důkazu o konvergentní, respektive divergentní validitě navrženého inovativního konceptu.

Bibliografická citace: Jelínek, M., Květon, P., & Vobořil, D. (2015). Innovative testing of spatial ability: Interactive responding and the use of complex stimuli material. *Cognitive Processing*, 16(1), 45–55.

STUDIE 8 Skryté aspekty v testování prostorové představivosti: Identifikace uplatňovaných stylů řešení položek

Jádrem prostorové představivosti jsou analogové kognitivní procesy, které fyzickou manipulaci s objekty zrcadlí v rámci mentální aktivity. V diagnostické praxi je však obtížné navrhnout takové položky, které by byly řešitelné pouze za pomoci analogových myšlenkových operací. Proto se v testech prostorových schopností můžeme setkat i s položkami, u kterých lze místo vizuálně-analogových procesů uplatnit verbálně-analytické strategie řešení. Z hlediska měření skutečné schopnosti prostorové představivosti jde o nežádoucí jev a v podstatě chybu měření. Například pokud má respondent dobrou prostorovou představivost, může si pohledem do mapy vytvořit mentální obraz krajiny a podle něj se při pochodu orientovat. V případě, že touto schopností nedisponuje v dostatečné

míře, může nedostatečnou orientaci v krajině kompenzovat vytvořením propozic definujících navzájem navazující jasně určené úseky. Z tohoto příkladu je zřejmé, že míra využití verbálně-analytických postupů je výrazně individuálně podmíněná. Proto se studie opírá o obecný metodologický přístup zaměřující se na osoby spíše než na proměnné, konkrétně byla využita technika latentních tříd, která umožňuje identifikovat latentní třídy osob vystihující významné trendy v datech. Pravděpodobnost určitého vzorce odpovědí je podle analýzy latentních tříd dána pravděpodobnostmi odpovědí na položky podmíněnými příslušností k jednotlivým latentním třídám a pravděpodobnostmi příslušnosti k latentním třídám. Analýza latentních tříd neumožňuje pouze popis jednotlivých typů na základě použitých indikátorů (položek), ale nabízí výzkumníkovi také podklady pro empiricky odůvodněnou úvahu o vhodném počtu tříd, což je umožněno tím, že analýza latentních tříd spadá mezi modelově zakotvené analytické přístupy zaměřené na určení povahy latentní proměnné, které bývají označovány souhrnně jako techniky analýzy latentní struktury.

Samotná analýza latentních tříd podobně jako jiné multivariační techniky (shluková analýza, faktorová analýza) nenabízí interpretaci nalezeného řešení, ale poskytuje výzkumníkovi prostor pro uplatnění úvah podložených teoretickými východisky. V případě této studie jsme význam nalezených latentních tříd vysoudili z charakteru testových úloh, tj. z hlediska míry možnosti uplatnění verbálně-analytických nebo vizuálně-analogových mentálních operací při jejich řešení. Využita byla data z Testu studijních předpokladů (N = 1540), který obsahuje subtest Prostorová představivost obsahující osm konceptuálně odlišných úloh. Z analýzy dat vyplynulo, že nejvhodnějším řešením je model uplatňující rozdělení osob do tří latentních tříd. Nejvíce zastoupená byla třída charakteristická relativně vysokou podmíněnou pravděpodobností správných odpovědí na všechny položky subtestu Prostorové představivosti. Další dvě třídy již byly odlišeny na základě profilu odpovědí na jednotlivé položky. V souladu s expertním posouzením charakteru položek bylo zjištěno, že při vzájemném porovnání jedna z těchto dvou tříd vykazuje vyšší pravděpodobnosti odpovědí na „verbálně-analytické“ položky (tj. položky, u kterých je možné ve zvýšené míře využívat verbálně-analytické strategie) a současně nižší pravděpodobnosti u „vizuálně-analogových“ položek. Výchozí model byl v souladu s tímto zjištěním modifikován pomocí fixací pravděpodobností odpovědí jednotlivých položek, přičemž modifikovaný model nevykazoval statisticky významně horší shodu oproti modelu bez fixací. Navržená interpretace jednotlivých latentních tříd na základě individuálních schopností uplatnění verbálně analytických strategií řešení byla úspěšně validizována na základě porovnání z hlediska výkonu osob v subtestu Analytické myšlení, o kterém lze oprávněně předpokládat, že je nejvýrazněji sycen právě verbálně-analytickými schopnostmi.

Tato studie ukázala, že celkový skóre testu prostorové představivosti může minimálně z části odrážet nejen (v principu) vizuálně-analogovou schopnost prostorové představivosti, ale také verbálně-analytickou schopnost, která v určitém typu položek může při řešení prostorovou představivost

víceméně nahradit. Tento fakt je dle našeho názoru nutno vzít v úvahu při interpretaci výsledků principiálně podobných testů.

Bibliografická citace: Jelínek, M., Květon, P., & Vobořil, D. (2013). Skryté aspekty v testování prostorové představivosti: Identifikace uplatňovaných stylů řešení položek. *Československá psychologie*, 57(4), 297–306.

STUDIE 9 Experimentální porovnání tachistoskopických měření: Reakce na Wundtovu kritiku

Tachistoskopická zařízení jsou velmi častou a téměř neodmyslitelnou součástí výbavy experimentálních laboratoří. Raní experimentální psychologové se nejvíce zajímali o základní principy a determinanty lidského vnímání a paměti, a právě tachistoskopy jako zařízení umožňující krátkodobou expozici podnětů ve standardizovaných podmínkách jsou spojeny se vznikem poznatků o rozpoznávání písmen, rozvoji percepční kapacity v průběhu životní dráhy, nebo charakteristikách vnímání pohybu na přelomu 19. a 20. století. Později v polovině 20. století využíval tachistoskop George Sperling ve svých klasických výzkumech, které vedly až k definici sensorického registru jako paměťového modulu odlišného od krátkodobé a dlouhodobé paměti. V současnosti se tachistoskopická prezentace podnětů používá v neurologických studiích a také ve studiích podprahového vnímání a studiích zaměřených na roli primingu v lidské percepci.

Ačkoli experimentální účel tachistoskopů zůstal neměnný po více než celé století, technologický pokrok vedl k rozdílným konstrukcím těchto zařízení. V rámci prezentované studie byl proveden srovnávací experiment, který byl zaměřen na čtyři typy tachistoskopických nástrojů představujících technické milníky základních konstrukčních variant - pádový tachistoskop, tachistoskop s fotografickým uzávěrem, počítačový tachistoskop s CRT obrazovkou, počítačový tachistoskop s LCD obrazovkou. Pomocí lineárního smíšeného modelu aplikovaného na data z vnitroskupinového experimentálního designu bylo sledováno, zda se jednotlivé typy tachistoskopů liší v celkové úrovni výkonu definovaného počtem správně reprodukováných stimulů a počtem chybně reprodukováných stimulů. Sledován byl také trend výkonu v průběhu času v závislosti na typu přístroje.

Z výsledků studie jednoznačně vyplývá, že typ tachistoskopické prezentace výrazně ovlivňuje výkon zkoumaných osob. Zatímco oba počítačem administrované postupy jsou v zásadě ekvivalentní, pádový tachistoskop a tachistoskop s fotografickým uzávěrem jsou z hlediska produkovaných výsledků odlišné jak od sebe navzájem, tak od počítačových tachistoskopů. Zatímco pádový

tachistoskop vykazuje od počítačem řízeného tachistoskopu odlišnost v celkovém výkonu pouze na počátku testování, přičemž s přibývajícimi pokusy se tento rozdíl eliminuje, u tachistoskopu s fotografickým uzávěrem je výkon zkoumaných osob výrazně horší než u všech ostatních instrumentů po celou dobu testování. Navíc tento tachistoskop vykazuje vyšší produkci chybných odpovědí respondentů. Zjištěné výsledky byly zasazeny do historického kontextu a byly diskutovány v souvislosti s Wundtem definovanými požadavky na validní tachistoskopickou prezentaci. Ve shodě s Wundtem definovanými kritérii adekvátní tachistoskopické situace předpokládáme, že absence fixačního bodu u tachistoskopu s fotografickou uzávěrkou (která tuto konstrukci odlišuje od ostatních přístupů) stojí v pozadí zhoršených výkonů. Navíc se domníváme, že také nežádoucí zvukový efekt doprovázející otevření a uzavření fotografické uzávěrky mohl výrazně interferovat s výkonem zkoumaných osob.

Bibliografická citace: Vobořil, D., Jelínek, M., & Květon, P. (2014). Experimental evaluation of tachistoscopic measurement: A step beyond Wundt's criticism. *American Journal of Psychology*, 127(2), 245–252.

4 ZÁVĚR

V oblasti psychologické diagnostiky bylo vytvořeno nezměrné množství nástrojů, které jsou mnohdy zaměřené na zachycení stejného konstruktů či jevu. I v případě, že výzkumník při plánování své studie dbá na výběr psychometricky kvalitních a ověřených nástrojů, vždy musí mít na paměti, že výsledky jsou závislé nejen na míře sledované charakteristiky respondentů, ale také na unikátních vlastnostech konkrétního použitého instrumentu. Prezentované studie zabývající se různými aspekty měření psychologických charakteristik navíc demonstrovaly, že zjištěné výsledky mohou být výrazně ovlivněny nežádoucími faktory. Teprve podrobná znalost konkrétního nástroje opřená o empirická zjištění umožňuje výzkumníkovi validně interpretovat získaná data v rámci jasně definovaných limitů a omezení.

SUMMARY

The presented habilitation thesis represents an annotated set of published studies generally focused on various approaches to psychological testing, such as item response theory, adaptive testing, innovative approaches to measuring specific capabilities or advanced statistical data processing techniques. The set of studies is thematically divided into two primary areas - methods and procedures used in personality assessment and methods and procedures in cognitive performance testing. In the case of personality assessment, research efforts aimed at the adaptation of widely used personality questionnaires grounded in the Big Five theory framework. In addition to the adaptation studies, the theoretical knowledge base in the field of measurement was enriched by studies exploring the possibilities of adaptive testing of personality characteristics, applying mathematical modeling to response strategies, as well as study proposing new statistical procedures for assessing the accuracy of measuring stereotypes in intercultural psychology. In the area of cognitive performance testing, three studies were devoted to spatial ability testing. One study described the process of development and psychometric evaluation of a multidimensional spatial reasoning test, the second one suggested novel, innovative item principles utilizing the advances of modern information technology, and the last one was investigating mental process involved in item responding strategies. The cognitive area ends with a study, which connects measuring cognitive abilities from the beginnings of scientific psychology to the present and focuses on the equivalence of measurement in the historical context.

PŘÍLOHY

- PŘÍLOHA A Hřebíčková, M., & Jelínek, M. (2019). NEO osobnostní inventář 3: Základní psychometrické charakteristiky české verze NEO-PI-3. *Československá psychologie*, 63(2), 160-176.
- PŘÍLOHA B Hřebíčková, M., Jelínek, M., Blatný, M., Brom, C., Burešová, I., Graf, S., ... Zábrodská, K. (2016). Big Five Inventory: Základní psychometrické charakteristiky české verze BFI-44 a BFI-10. *Československá psychologie*, 60(6), 567–583.
- PŘÍLOHA C Jelínek, M., Květon, P., & Vobořil, D. (2011). Adaptivní administrace NEO PI-R: Výhody a omezení. *Československá psychologie*, 55(1), 69–81.
- PŘÍLOHA D Jelínek, M., Květon, P., & Vobořil, D. (2010). Assessment of response pattern aberrancy in Eysenck Personality Inventory. *Annales Psychologici*, 58(14), 37–46.
- PŘÍLOHA E Hřebíčková, M., Möttus, R., Graf, S., Jelínek, M., & Realo, A. (2018). How accurate are national stereotypes? A test of different methodological approaches. *European Journal of Personality*, 32(2), 87–99.
- PŘÍLOHA F Květon, P., Jelínek, M., & Vobořil, D. (2014). Testing of spatial ability: Construction and evaluation of a new instrument. *Studia Psychologica*, 56(3), 233–251.
- PŘÍLOHA G Jelínek, M., Květon, P., & Vobořil, D. (2015). Innovative testing of spatial ability: Interactive responding and the use of complex stimuli material. *Cognitive Processing*, 16(1), 45–55.
- PŘÍLOHA H Jelínek, M., Květon, P., & Vobořil, D. (2013). Skryté aspekty v testování prostorové představivosti: Identifikace uplatňovaných stylů řešení položek. *Československá psychologie*, 57(4), 297–306.
- PŘÍLOHA I Vobořil, D., Jelínek, M., & Květon, P. (2014). Experimental evaluation of tachistoscopic measurement: A step beyond Wundt's criticism. *American Journal of Psychology*, 127(2), 245–252.