

Diskuze

VALIDITA MERANIA: HĽADANIE DÔVODOV ABSENCIE KONSENZU

IVAN ROPOVIK

Pedagogická fakulta PU, Prešov

ABSTRACT

Validity of measurement: exploring reasons for the absence of general consensus

I. Ropovik

The article discusses the concept of validity and explores the reasons for the absence of a widespread consensus in its definition and perception. Applying a diachronic perspective, the most influential conceptions of validity are being confronted with an emphasis on both conceptual and epistemological issues. Based on that analysis, it is argued that differences between several conceptions of validity are not just a terminological quibble. Furthermore, it is demonstrated that the meaning of even the simplest and mostly cited definition (a test is valid if it measures what it

purports to measure) depends on some choices that are in fact arbitrary. These choices made further imply the epistemological procedures needed for the attribution of the “validity” label and whether validity should be considered a relational or a causal concept.

key words:

validity,
consensus,
measurement,
causality

klúčové slová:

validita,
konsenzus,
meranie,
kauzalita

1 ÚVOD

S ústupom psychofyziky vzniká v prvých dekádach dvadsiateho storočia v psychológii potreba nového konceptu zvaného validita. Z predtým redundantného (a v mnohých exaktných vedách prakticky neexistujúceho) konceptu sa postupne vyvinie ústredný pojem metodológie psychologickéj vedy. Čo je však dôvodom redundantnosti konceptu validity v exaktných vedách a na druhej strane niekedy možno až mystifikovanému, no prominentnému významu v psychológii a iných sociálnych vedách? Dôvodov je možné nájsť viacero. Psychológia skúma psychické atribúty, pričom tie sú značne komplexné a nie je ich navyše možné priamo merať. Na druhej strane, aj v rámci viacerých exaktných vied existujú multideterminované premenné, ktoré taktiež majú často latentný charakter. Rozdielom avšak je, že psychológia nedisponuje obdobným experimentálnym aparátom ako exaktné vedy a možnosti izolácie a manipulovateľnosti premenných tak nie sú na dostatočnej úrovni (viď Trendler, 2009).

Došlo: 18. 12. 2013; I. R., Katedra predškolskej a elementárnej pedagogiky a psychologie, Pedagogická fakulta PU, Ul. 17. novembra 15, 080 01 Prešov, Slovenská republika; e-mail: ivan.ropovik@unipo.sk

Príspevok je výstupom riešenia projektu APVV-0281-11 (Exekutívne funkcie ako štrukturálny komponent schopnosti učiť sa: diagnostika a stimulácia. Zodpovedný riešiteľ Iveta Kovalčíková).

Myseľ človeka má navyše dynamický a výrazne reaktívny charakter, čo značne komplikuje akékoľvek meranie. Prechod od pomerne jednoduchého psychofyzikálneho merania na úroveň komplexných a problematicky merateľných psychických atribútov v kombinácii so snahou o zotrvanie v pozitivistickej orientácii tak celkom logicky determinoval nevyhnutnú potrebu vzniku konceptu validity. Úlohou validity v čase jej vzniku teda bolo vysporiadanie sa s novovzniknutými nejasnosťami a problémami merania, čo v konečnom dôsledku umožnilo psychológii vyhnúť sa eventuálnemu označeniu za metafyzickú disciplínu.

V záujme približovania sa štandardom exaktných vied je potrebné, aby validita ako jeden z ústredných pojmov psychometriky mala jednoznačnú interpretáciu s jasne vyvoditeľnými empirickými implikáciami. Očakávanie všeobecne akceptovaného konsenzu v otázke jej definovania a záberu obsiahnutých konceptov by však bolo mylné. Existujú totiž podstatné rozpory v percepcii validity medzi vedeckou praxou a teóriou validity (viď Newton, Shaw, 2013), pričom aj v rámci samotnej teórie validity existujú výrazne odlišné koncepcie (Anastasi, 1950; Borsboom et al., 2004; Cronbach, Meehl, 1955; Kane, 2006; Messick, 1989), ktoré sa nezhodujú dokonca ani v tom, či sú otázky merania pre koncept validity relevantné. Majú však tieto rozpory reálne dôsledky a sú teda viac ako iba hry postavené na terminologickej manipulácii? Kam až siahajú korene absencie skutočného konsenzu?

2 VÝVOJ KONCEPTU VALIDITY

2.1 Kritériálny model

Historicky prvotná koncepcia validity sa vzťahovala špecificky na meranie a znela, že test je validný, ak meria to, čo má merať (napr. Kelley, 1927). Dobový kontext tejto definície podliehal filozofii logického pozitivizmu, ktorý striktne izoloval pozorovateľné aspekty reality od „neskúmateľných a metafyzických“ teoretických konštruktov (Carnap, 1936), pričom zo zásady popieral ich existenciu. Účelom testovania bolo čo najpresnejšie stanoviť hodnotu cieľovej premennej u testovaného subjektu, pričom úlohou validity bolo prisúdiť tomuto odhadu pravdivostnú hodnotu. Mechanizmus dokladovania validity tu primárne pozostával z dvoch synergicky fungujúcich komponentov.

Prvým komponentom bolo kritérium. Ak boli výsledky testovania v tesnom vzťahu k stanovenému kritériu meraného konštrukt, test bol považovaný za validný. Problémom však bolo, že v psychológii je spravidla veľmi náročné identifikovať pozorovateľné kritérium, ktoré by bolo v tesnom vzťahu s teoretickým konštruktom. Na druhej strane, vedomosť o nedokonalosti kritéria sama o sebe implikuje otázku validity kritéria samotného, čo pri kritériálnom poňatí validity teoreticky vedie k nekonečnému regresu. Na riešenie tohto problému slúži druhý (z pozitivizmu prameniacy) komponent, a tým je paradigma operacionalizmu. Adaptáciou myšlienok fyzika Percyho Bridgmana (1927) vznikla napohľad elegantná doktrína, ktorá našla široké uplatnenie a to vďaka jej schopnosti stotožniť inak latentný konštrukt so samotnou operáciou jeho merania. Táto jednoduchá sémantická rovnica redukujúca teoretický konštrukt na procedúru merania tak umožňuje výstupu merania priradiť status samotného teoretického konštrukt. Operačné definovanie validity meraného kritéria psychológii poskytlo „pevný bod“, ktorý bol potrebný pre pozitivizmom vyžadovanú kompletnú elimináciu čisto teoretických konštruktov (viď Bickhard, 2001). Koncepcia validity testu postavená na tesnosti vzťahu s externým kritériom dominovala v psychológii až do začiatku druhej polovice 20. storočia (Zumbo, 2009), pričom túto dominanciu dokladuje vtedajší striktne empiristický konsenzus, že „test môže objektivne validizovať

jedine meranie špecificky definovaného kritéria. Tvrdenie, že test je mierou čohokoľvek iného než daného kritéria, je čírou špekuláciou“ (Anastasi, 1950, s. 67).

Kriteriálny model validity je v súčasnosti užitočným nástrojom najmä v aplikovaných doménach (t.j. prvoradé je dosiahnuť žiadanú funkcionálnu psychologickejho nástroja a nie dokladovanie jeho meracích schopností), kde je ľahko možné nájsť validné kritérium. Avšak v mnohých iných kontextoch nebolo problémom ani tak rozhodnúť o samotnej validite kritéria, problémom bolo nájsť akékoľvek kritérium (Cronbach, Meehl, 1955). Neúspech pri hľadaní takýchto kritérií (a zároveň neexistencia obsahového definičného oboru skúmaných konštruktov) vytvoril dopyt po inom, alternatívnom modeli validizácie psychologických testov. Hoci bol model kritériálnej validity úspešný v tom, že viedol k rozvoju viacerých psychologických teórií, to paradoxne viedlo k jeho úpadku, keďže tieto teórie už ďalej nebolo možné validizovať len za pomoci kritéria, či predikcie (Strauss, Smith, 2009).

2.2 Cronbach

Na začiatku päťdesiatych rokov minulého storočia bola v rámci jednej z komisií APA iniciovaná snaha o reformuláciu konceptu validity, ktorý by tak viac vyhovoval potrebám testovania v rámci poradenskej a klinickej praxe (Kane, 2001). Z tejto snahy rezultovalo vydanie prvých Štandardov, ktoré už popri obsahovom a kritériálnom modeli validity zahŕňali aj nový koncept konštruktivej validity. Tá predstavovala nadradený (no zatiaľ nie unitárny) koncept zjednocujúci aspekty obsahovej ako aj kritériálnej (súbežnej a prediktívnej) validity, ktorého primárnym cieľom avšak bolo ukotviť skúmaný konštrukt v širšom teoretickom systéme. Konštruktová validita sa tak stala vyjadrením potreby skôr explanačného, ako ad hoc relačného charakteru validity. Explanácia validity konštruktov vyžadovala v prvom rade postulovanie “existenčných” pravidiel, v rámci ktorých sa daný konštrukt vyskytuje (tzv. nomologickej siete), následné definovanie manifestácií definovaných konštruktov (zostavenie korešpondujúcej siete pozorovaných premenných) a nakoniec konfrontáciu výsledkov empirického testovania operacionalizovaných hypotéz so systémom vzťahov predikovaných nomologickou sieťou (Cronbach, Meehl, 1955).

Keďže v rámci modelu kritériálnej validity bol význam konštruktov jasne definovaný kritériom (pričom absencia kritéria viedla k nedefinovateľnosti významu), odstránenie požiadavky na existenciu kritéria identického s konštruktom, ktorý je meraný testom, psychológii otvorilo široké možnosti skúmania psychologických konštruktov, ktoré boli z dôvodu absencie jasnej obsahovej domény alebo kritéria nevalidizovateľné v rámci predošlého modelu validity. Model konštruktivej validity bol pritom napriek snahe o skúmanie teoretických konštruktov stále v súlade s logickým pozitivizmom, keďže relačné definovanie významu teoretického konštruktov prostredníctvom nomologickej siete iných teoretických konštruktov nijakým spôsobom nutne nevyžaduje priradenie ontologického statusu jej elementom (Colliver et al., 2012). Rovnako ako v rámci modelu kritériálnej validity, interpretácia merania pozorovateľných premenných ako meranie čisto teoretických konštruktov je sémanticky legítimná, keďže nevyžaduje postulovanie ich existencie. Idea nomologickej siete a jej začlenenie do štandardov psychologickej vedy tak v nasledujúcich rokoch výrazne oslabilo dovtedy dominantný empirizmus, keďže pozorované premenné stratili svoju exkluzivitu v determinácii konštruktov a stali sa iba indikátormi.

Koncept nomologickej siete má avšak, napriek sľubnému potenciálu prinavrátiť explanáciu do centra psychologickej vedy, niekoľko problémov. Po prvé, v kontraste s často ateoretickým kritériálnym modelom, konštrukty sú ukotvené iba ich významom v rámci širšej teórie (Kane, 2001). To so sebou avšak prinieslo problém empi-

rickej „pod-identifikovanosti“, t.j. z jednej známej (sady skórov) je potrebné vyvodit' dve neznáme (validitu nástrojov a zároveň správnosť teórie). Ak všetky premenné v nomologickej sieti závisia na teórii, tak akékoľvek empirické testovanie musí vo-pred predpokladať validitu teórie samotnej. Možnosť testovať teóriu teda nutne vyžaduje, aby aspoň niektoré premenné neboli závislé na žiadnej teórii. Pri absencii takýchto premenných je v konečnom dôsledku potrebné nahradiť potrebné informácie postulovaním predpokladov (Coombs, 1964) a axiomaticky definovať určité elementy alebo vzťahy v rámci nomologickej siete. Bez toho by nebolo možné určiť, či je dôvodom eventuálnej absencie zhody teórie a dát nevalidita skúmaného nástroja, nevalidita nástrojov merajúcich explanačné konštrukty, neplatnosť teórie, alebo kombinácia týchto dôvodov.

Druhý problém sa týka samotnej implementácie modelu konštruktivej validity. Už pri formulácii konceptu konštruktivej validity Cronbach a Meehl (1955) konštatujú prílišnú vágnosť vtedajších psychologických teórií (nomologických sietí), čo v danej dobe neumožňovalo pevné uchopenie psychologických konštruktov, keďže podľa nich sa ešte len „*nachádzame v procese objavovania [psychologických] pravidiel*“ (s. 292). S konštruktovou validitou sa avšak bude v nasledujúcich rokoch „*zaoberať ako s odpadkovým košom...kde kapitoly s týmto názvom sú v testových manuáloch väčšinou neusporiadanou zbierkou korelácií s inými rôznorodými testami a demografickými premennými...a to bez súbernej argumentácie*“ (Cronbach, 1980, s. 44).

Po tridsiatich rokoch je Cronbach (1988) nútený konštatovať praktickú neexistenciu nomologických sietí. To, čo sa pôvodne malo stať štandardom vedeckej praxe, už Cronbach prezentuje ako idealizáciu a označuje to ako „silný program“. Naproti tomu pomenúva v praxi aplikované postupy konštruktivej validizácie ako „slabý program“, ktorý už je výrazne vzdialený od vtedajšej pozitivistickej pozície vyžadujúcej prísnu formalizáciu skúmaných teórií. Slabý program tak Cronbach označuje za „*čirý exploračný empirizmus*“, pričom kritizuje praktickú absenciu formálnych teórií konštruktov a konfirmacionistické „*vítanie akejkoľvek korelácie*“ (Cronbach, 1988, s. 12). Nový, slabý program validizácie nemá jasné ciele, je eklektickým a pravdepodobne nikdy nekončiacim procesom so značným potenciálom pre oportunistiku vo voľbe dôkazov, procesom založeným často iba na racionalizácii a posteriori (Kane, 2001; Sijtsma, 2009).

Perspektíva Cronbacha a Meehla dominovala konceptu validity v prvých štyroch vydaniach štandardov, ktoré sa líšili primárne iba v terminológii, pričom zmeny reflektovali skôr jemné korekcie v koncepte validity. Pod vplyvom dávnejšie formulovanej idey, že „*keďže prediktívna, súbežná a obsahová validita sú v podstate ad hoc, konštruktová validita je z vedeckého hľadiska validitou v celom rozsahu*“ (Loevinger, 1957, s. 636), sa validita explicitne unitarizovala pod hlavičkou konštruktivej validity, ktorá už zahŕňala aj obsahové (logické) a kriteriálne (empirické) zdroje dôkazov validity. Cieľom pritom bolo redukovat' praktizovaný eklektizmus vo voľbe dôkazov validity (Rogers, 1995). Na čo však psychológia odpovedať nevedela, bola absencia nomologických sietí (formálnych teórií meraných konštruktov alebo atribútov), splňajúcich predstavy pozitivizmu (viď Hempel, 1965).

2.3 Messick

Zásadnou ideovou zmenou čiastočne reflektujúcou už dlhšiu dobu silnejúce post-pozitivistické tendencie v sociálnych vedách bola až koncepcia Samuela Messicka (1989). Tá by sa dala zhrnúť do nasledujúcich piatich bodov. (1) validita nie je vlastnosťou samotného testu, ale interpretácie výsledkov, je teda možné validizovať iba samotné interpretácie a spôsoby použitia testového skóre; (2) konštruktová validita už nie je typom validity, ale súhrnnou konceptualizáciou validity; (3) validita nie je binárnym

úsudkom, existuje ako kontinuum vyjadrené v stupňoch; (4) validita je kontextuálne a hodnotovo podmienenou formou úsudku; (5) vzhľadom na existenciu množstva rozdielnych populácií, podmienok, účelov a dôsledkov testovania, definitívne rozhodnutie o validite môže mať prinajlepšom iba asymptotický charakter. Messick odstránil z konceptu validity dôraz na striktnú formalizáciu psychologických teórií (Kane, 2001) a validitu konceptualizoval ako úsudok založený nie len na dôkazoch, ale aj hodnotách výskumníka (kladením dôrazu na dôsledky testovania), čím implicitne vyjadřil pochybnosť, či je vôbec možné dosiahnuť objektivnosť validity.

Messick tak definuje validitu nasledujúco. „*Validita je súhrnným hodnotiacim úsudkom, do akej miery empirické dôkazy a teoretické zdôvodnenia podporujú adekvátnosť a primeranosť dedukcií a konaní založených na testovom skóre a iných spôsoboch hodnotenia....Vo všeobecnosti, validita je induktívnym súhrnom existujúcich dôkazov pre interpretáciu a použitie testového skóre ako aj možných dôsledkov tejto interpretácie a použitia*” (Messick, 1989, s. 13). Aktuálne Štandardy do značnej miery stavajú na Messickovej konceptualizácii a definujú validitu ako mieru, v akej empirické dôkazy a teórie podporujú interpretáciu testového skóre pri doporučenom použití testu (AERA, APA, NCME, 1999, s. 9). V porovnaní s Cronbachovou prvotnou konceptualizáciou (rovnako vid' Anastasi, 1950; Cureton, 1951), ktorý definoval validitu ako „*mieru, do akej vieme, čo test meria alebo predikuje*” (Cronbach, 1949, s.48)¹⁾, tak zahŕňa jeden podstatný obrat. Podľa aktuálnej konceptualizácie sa validita nevzťahuje na test, ale na interpretáciu testového skóre (Kane, 2006; Messick, 1989), čo je snahou o explicitné vyjadrenie faktu, že výsledky testovania nie je možné interpretovať bez zohľadnenia kontextuálnych aspektov (populácia, podmienky testovania).

V pozadí posunu v asociovaní validity (validita testu vs. validita interpretácií) tak z definície validity logicky vypadol koncept merania. Na rozdiel od otázky, čo meria test, totiž nedáva zmysel pýtať sa, čo meria interpretácia. Prvotná konceptualizácia validity ako vlastnosti merania (resp. procedúry merania – testu) sa tak postupne pod vplyvom post-positivistického diskurzu transformovala na vlastnosť vzťahujúcu sa na interpretáciu a použitie testového skóre k istému účelu. Otázka či a čo test meria, tu už nie alfou a omegou validity. Validita je tak v Messickovom ponímaní výrazne širším a permissívnejším konceptom syntetizujúcim viaceré prvky ostatných existujúcich koncepcií validity, pričom umožňuje aj interpretácie (vzťahujúce sa na užitočnosť, či dôsledky testovania), ktoré vôbec nie sú založené na meraní (Borsboom et al., 2009).

2.4 Borsboom

V posledných rokoch silnejú alternatívne tendencie formulovať validitu ako užší, technickejší pojem (vid' Newton, Shaw, 2013), ktorý sa vracia k prvotnej definícii, keď validita bola vlastnosťou testu merať to, čo má merať (Kelley, 1927). V súčasnosti azda teoreticky najkoherentnejším prístupom reprezentujúcim túto líniu je Borsboomova koncepcia validity (Borsboom et al., 2004). Táto koncepcia je zároveň v najvýraznejšej opozícii k Messickovmu poňatiu validity, keďže je vo faktickom rozpore so všetkými piatimi, vyššie spomenutými bodmi Messickovej koncepcie (vid' Borsboom et al., 2004; Borsboom et al., 2009). Pre Borsbooma je test validným meraním istého atribútu vtedy a iba vtedy, ak (1) daný atribút existuje a zároveň (2) variácia v atribúte kauzálne produkuje variáciu v testovom skóre. Ak teda niečo neexistuje, nie je možné to merať. Ak to existuje, ale nemá to vplyv na výsledok merania, nejedná sa o meraný atribút (Borsboom et al., 2004). Keďže neexistujúce entity logicky nemôžu

¹⁾ V skutočnosti sa Cronbach v ďalších prácach vyhýbal explicitnému definovaniu validity, neskôr avšak taktiež hovorí o validizovaní interpretácií (1955).

mať akékoľvek kauzálne efekty, prvý bod je v skutočnosti redundantnou podmienkou, keďže tá je zastúpená v bode druhom. Cieľom však bolo explicitné zasadenie konceptu validity do filozofickej paradigmy realizmu. Keďže validita je v rámci tejto koncepcie kauzálnym a nie relačným konceptom, realizmus vo vzťahu k existencii meraných atribútov je jedinou možnou cestou, ak nechceme poprieť základy súčasnej vedy formulované v rámci existujúcich teórií kauzality (viď Bunge, 2008; Michell, 2005; Pearl, 2009).

Naproti tomu Messick svoju koncepciu explicitne zasadzuje do filozofie konštruktívneho realizmu (viď Loevinger, 1957; Messick, 1989), čím vyjadruje postoj, že konštrukty skúmané v psychológii môžu a nemusia zastupovať reálne existujúce mentálne atribúty. Messickov primárny nástroj poznania validity (nomologická sieť) tak nevyžaduje dôkazy existencie meraných konštruktov pre atribúciu validity, čo činí Messickovu koncepciu z hľadiska dokladovania validity merania výrazne permisívnejšou (Hood, 2009). Napriek deklarovanej konštruktívno-realistickej pozícii avšak Messickom preferované postupy relačného charakteru nedokážu rozlíšiť teoretické koncepty, ktoré iba sumarizujú pozorovanú realitu, od reálne existujúcich atribútov.

Validita v rámci Borsboomovej realistickej koncepcie (na rozdiel od predošlých konceptualizácií) nie je kontinuum, ale konceptom nadobúdajúcim iba hodnoty pravda-nepravda. Totiž to, či daný atribút vplýva na výsledok merania, je podľa Borsbooma žiaduce zodpovedať prostým áno alebo nie (Borsboom et al., 2004). To, že nástroj meria to, čo má merať, avšak neznamená, že nástroj je kvalitný a má optimálne meracie schopnosti²⁾. Validita v binárnom poňatí je kvalitou nezávislou od ďalších vlastností merania ako je reliabilita, invariancia, užitočnosť pre účely predikcie, pokrytie obsahovej domény, atď. Táto pozícia je v príkrom rozpore s inak intuitívnu ideou, že validita je kvantitatívnym konceptom existujúcim v rôznych stupňoch. Napriek širokej akceptácii konceptu stupňov validity psychometrika nedisponuje všeobecnou formálnou definíciou ani procedúrou ich odhadu. Nie je totiž jasné, funkciou ktorých vlastností merania by stupne validity mali byť, aká by mala byť váha týchto jednotlivých vlastností v rámci tejto funkcie, ako a vôbec či je možné porovnávať stupne validity rôznych nástrojov (aký je napr. rozdiel v stupňoch validity medzi ľubovoľným subtestom WISC, mikrometrom a termostatom?), atď. (viď Markus, Borsboom, 2013).

Žiaden z týchto problémov nie je problémom napr. vo formálne definovanom koncepte (stupňov) reliability, kde tá je definovaná ako ρ^2_{XTP} , t.j. štvorec korelácie medzi pozorovaným a pravým skóre v populácii (Lord, Novick, 1968). Zároveň existujú viaceré metódy odhadu tejto hodnoty a hoci sú viaceré z nich podrobované kritike, sú tieto metódy jasne formálne definované, čím vzniká predpoklad presnejšieho a objektívnejšieho určenia proporcie systematickej, reliabilnej variance v pozorovanom skóre. Borsboomova koncepcia validity ako základu a nutnej podmienky kvalitného merania tak umožňuje postuláty typu (1) nástroj *meria* nepresne (test okrem cieľového atribútu meria aj iné, nežiaduce premenné), alebo nástroj *meria* v rôznych populáciách inak (nie je zachovaná invariancia merania)³⁾. Ak test *meria* daný atribút, je

²⁾ Zároveň, ak test *nemeria* daný atribút, neznamená to, že nie je validným. Test stále môže byť napríklad validným pre účely predikcie (alebo klasifikácie do skupín, či pokrytia obsahovej domény), keďže optimalizácia meracích a prediktívnych kvalít nástroja sú divergentnými líniami konštrukcie testu. V takomto prípade avšak nie je možné ale ani nutné interpretovať testové skóre ako výstup merania.

³⁾ Neplatí, že $f(X|\theta)=f(X|\theta,P)$, kde f je funkciou rozdelenia hustoty pravdepodobnosti, X je pozorované skóre, θ je latentnou premennou a P je indexom príslušnosti k istej populácii. Test je v Borsboomovej koncepcii validným aj pri nerovnosti oboch funkcií, t.j. ak je okrem cieľového atribútu citlivý aj na príslušnosť k skupine.

validný, neznamená to však, že test je nutne kvalitný a výsledok interpretovateľný. To, či test meria presne, či je dodržaná invariancia merania pre rôzne populácie, či test dokáže predikovať kritérium, alebo či pokrýva obsahovú doménu konštruktú, spadá pod gesciu iných psychometrických konceptov. Napr. nízka reliabilita pre Borsbooma neimplikuje nízku validitu (vid' Markus, Borsboom, 2013). Meranie IQ hoci aj generátorom náhodných čísel nie je nevalidným preto, lebo má nulovú reliabilitu (tá je irelevantná), ale preto, lebo IQ nemá vplyv na takúto „meranie“. V prvom rade je potrebné, aby test preukázateľne zachytával kauzálny vplyv daného atribútu, až potom má zmysel sústrediť sa na optimalizáciu jeho meracích kvalít (napr. redukciou vplyvu nežiaducich premenných). To, či je interpretácia opodstatnená, aké je použitie a aké dôsledky z testovania vyplývajú, sú teda samostatnými, nadradenými otázkami k otázke validity.

Dalším zásadným rozdielom je, že aktuálna Messickova koncepcia formuluje validitu ako hodnotovo podmienený úsudok a nedefinuje, aké dôkazy oprávňujú prehlásiť interpretáciu testu ako vysoko validnú (Hood, 2009). Messick pri rôznych metódach síce konštatuje, že dané metódy nie sú nikdy dostatočujúce pre preukázanie konštruktivej validity (Messick, 1989), no žiadne dostatočné podmienky (ani len teoreticky ideálne) neuvádza. Naproti tomu Borsboom explicitne definuje dve dostatočné podmienky ako aj potrebný druh dôkazov, t.j. poznanie kauzálneho modelu, resp. procesov vedúcich k testovej odpovedi. Validita tu tak nie je niečím neuchopiteľným, ale je prostým poznaním príčin (Zumbo, 2009).

Proti Borsboomovmu prístupu je avšak možné formulovať viaceré námietky. Po prvé, Borsboomov prístup oproti Messickovmu nie je natoľko špecifický ohľadom postupov dokladovania validity. Zatiaľ čo Messick vníma validitu ako epistemologický koncept (Messick hovorí primárne o validizácii, keďže validitu chápe ako výsledok validizácie) a ponúka rozsiahly zoznam široko aplikovateľných metód jej dokazovania, Borsboom sa primárne zaoberá ontológiou a nie epistemológiou, čo z pragmatického hľadiska znižuje užitočnosť Borsboomovej koncepcie. Messick navyše na rozdiel od Borsbooma nepotrebuje vo svojej koncepcii (z pohľadu pozitivizmu problematiku) postulovanie existencie psychických atribútov. Po druhé, výraznou prekážkou pre širšie uplatnenie Borsboomovej koncepcie je komplexnosť mnohých psychologických konštruktov a v stave súčasného poznania je len ťažko možné očakávať konštrukciu kauzálnych modelov (snáď s výnimkou elementárnejších kognitívnych funkcií), vysvetľujúcich ako rôzne úrovne skúmaných konštruktov vedú k rôznym testovým odpovediam (Sijtsma, 2009). Po tretie, pojem validity u Borsbooma stráca akúkoľvek súvislosť s pojmom kvalita, vďaka čomu môže byť v limitných prípadoch atribúcia validity z pragmatického hľadiska absolútne bezpredmetná. Ak platí, že skutočná náhodnosť je absenciou kauzality a reliabilita je v zásade absenciou náhodnosti, tak nemusí byť vhodné formulovať validitu ako vlastnosť nezávislú na reliabilite. V rámci Borsboomovej koncepcie totiž aj triviálna, no nenulová proporcia pravej variancie so známym kauzálnym pôvodom stačí na atribúciu validity testu, ktorý má inak prakticky nulovú meráciu alebo prediktívnu schopnosť.

3 ARBITRÁRNE VOLBY V CHÁPANÍ VALIDITY

Má teda vzhľadom na viaceré kategoriálne rozdiely diskutovaných koncepcií vôbec zmysel hľadať univerzálnu definíciu validity? Niektorí autori vyjadrujú názor, že vyhliadky na dosiahnutie univerzálneho konsenzu sú slabé (napr. Newton, Shaw, 2013). História konceptu validity ukazuje, že definovanie validity ako stále širšieho, všeobecne platného pojmu, ktorý integruje viaceré, do značnej miery nezávislé kategórie (meranie, opodstatnenosť interpretácie, použitie testového skóre a dôsledky

testovania), do jediného úsudku, nie je plodnou cestou, a to z hľadiska teoretickej ako aj epistemickej neukotvenosti takéhoto prístupu. Hoci užšia, technickejšia konceptualizácia validity, vzťahujúca sa výlučne na meranie, tieto problémy výrazne redukuje, ani ona nie je jednoznačne uchopiteľná. Pre ilustráciu si vezmime v praxi najbežnejšie uvádzanú definíciu, že test je validný, ak meria to, čo má merať⁴⁾. Povedzme, že túto definíciu aplikujeme v kontexte merania konkrétneho psychologického konštraktu, napríklad pracovnej pamäti (Working memory – WM) konkrétnym elementárnym testom, napríklad testom Spätné opakovanie čísel (Digit span backward – DSB). DSB vyžaduje zapamätanie série čísel, pričom zopakovať ich je potrebné v inverznom poradí. WM zvykne byť definovaná ako aktívny pamäťový systém zodpovedný za dočasné uchovanie a simultánne spracovanie informácií (viď Baddelley, 2000). DSB je teda validný, ak platí výrok „DSB meria WM“. Logicky nasledujúcim krokom po jasnej definícii podmienky validity by mala byť fáza overenia tohto výroku (validizácia), vyžadujúca voľbu konkrétnych nástrojov testovania zhody teórie a dát ako aj stanovenie formálneho psychometrického modelu. Nie je tomu tak. Tento zdanlivo jasný výrok totiž v skutočnosti vôbec nie je jednoznačný. Existuje totiž viacero interpretácií, pričom voľba interpretačného rámca značne ovplyvňuje možnosti a voľby postupov dokladovania validity. Ešte závažnejší je však fakt, že na základe tej istej sady dôkazov je možné v rôznych interpretačných rámcoch prisúdiť tomuto výroku rozdielnu pravdivostnú hodnotu.

Interpretácia samotného výroku „DSB meria WM“, ako aj relevancia možných dôkazov, variujú v závislosti od odpovedí na dve otázky, a to (1) či je WM reálne existujúcim atribútom a (2) ako interpretujeme slovo „merať“ (Markus, Borsboom, 2013; Michell, 2011). Existencia je vo všeobecnosti dichotomickou záležitosťou. Niečo buď je, alebo nie je. Pritom iba to, čo reálne existuje, môže spôsobovať pozorovateľné, reálne rozdiely medzi ľuďmi. Existencia atribútu ako WM je teda základnou podmienkou platnosti výroku „DSB meria WM“. Jedinou alternatívnou možnosťou k existencii je neexistencia, t.j. atribúty sú iba konštrukciou. Je však možné, aby zároveň platilo, že „DSB meria WM“ a „WM neexistuje“ (Borsboom et al., 2004), alebo aby neexistujúci atribút (konštrukt) spôsoboval rozdiely medzi ľuďmi pozorované v rámci akejkoľvek paradigmy merania? Táto logická nekonzistentnosť striktne konštruktivistickej pozície nakoniec viedla k jej nahradeniu realizmom (Borsboom et al., 2004; Hood, 2009; Markus, Borsboom, 2013; Michell, 2005) alebo konštruktívnym realizmom, ktorý vraví, že konštrukt ako taký neexistuje ako samostatná entita, ale jedná sa o model zastupujúci mikrosystém neznámych, synergicky pôsobiacich atribútov (Messick, 1989). Rozdiel medzi aktuálnou implementáciou týchto koncepcií spočíva primárne v dôraze na formalizáciu a z toho prameniaca potrebu dôkazov o existencii konštruktov. S výnimkou pragmaticky orientovaných koncepcií (napr. Kane, 2006), ktoré sa týmito otázkami nezaoberajú, rôzne formy realizmu vo vzťahu k existencii atribútov majú v súčasnosti silnú pozíciu a formujú súčasné vnímanie validity.

Existencia, resp. neexistencia atribútu podmieňovala definovateľnosť výroku validity. Druhou otázkou je, ako interpretujeme slovo „merať“. Táto interpretácia už avšak podmieňuje samotnú pravdivosť daného výroku. Všeobecne v rámci vedy (tzv. klasická teória merania), meranie je definované ako určenie pomeru veľkosti atribútu k inej veľkosti (jednotke) tohto atribútu. Avšak na to, aby vzťah medzi dvoma rôznymi veľkosťami atribútu ostal invariantný vzhľadom na voľbu jednotky, daný atribút

⁴⁾ Hovoriť o validite testu je v rozpore z Messickovou koncepciou validity ako vlastnosti interpretácií (1989). Ak však v rámci oboch diskurzov akceptujeme, že fungovanie testu je závislé na relevantných vlastnostiach testovanej populácie ako aj okolnostiach testovania, stáva sa tento rozpor už iba slovnou hrou bez vecného obsahu.

musí mať pre zachovanie tejto invariance istú internú štruktúru (viď Michell, 2005). Ak je táto interná štruktúra v súlade so základnými axiómami merania (Hölder, 1901), postavenými na princípoch formulovaných už Aristotelom a Euklidom, môže sa daný atribút označiť za kvantitatívny. Úroveň, do akej sa meraniu následne darí produkovať skóre, ktoré adekvátne reflektuje (kvantitatívnu) štruktúru premennej, sa nazýva škálou. Z dôvodov nejasností, či psychické atribúty sú kvantitatívne (viď Urbánek et al., 2012), a absencie stálych jednotiek (kvôli značnej arbitrárnosti a výrazne nedokonalnej reprezentácii kvantitatívnej štruktúry atribútov škálami), meranie v rámci tejto všeobecne platnej koncepcie merania nie je možné (viď Barrett, 2005; Markus, Borsboom, 2013; McGrath, 2005; Michell, 2005) a psychologické testy teda bez splnenia axióm klasickej teórie merania nemôžu byť v rámci tejto pozície považované za validné. Klasická teória merania teda vyžaduje znalosť deterministického kauzálneho vzťahu medzi príčinou (atribútom) a následkom (výstupom merania), čo je v psychológii z empirického hľadiska spravidla neuchopiteľné a to z dôvodu absencie deterministických formálnych teórií meraných atribútov.

Takéto závery boli nezlučiteľné s pozitivistickými ideálmi psychológie, čo logicky vytváralo dopyt po inej, v psychológii široko uplatniteľnej, teórii merania. Napriek novej výhrade, či oprávňuje skúmanie rôznych prirodzených systémov aplikáciu rôznych teórií merania, takúto teóriu formuloval Stanley S. Stevens. Meranie tu definuje na tzv. reprezentacionalistickom princípe, ako „*priradzovanie číslíc objektom alebo javom podľa pravidiel*“, pričom rôzne pravidlá viedli k rôznym druhom škál (Stevens, 1946, s. 677). Stevens nešpecifikuje konkrétne pravidlá, ale iba definuje škály na základe prípustných matematických operácií (transformácií) zachovávajúcich formu danej škály. Stevensova definícia „*redukuje problém, čo je a čo nie je meranie, na jednoduchú otázku... Ak dokážeme stanoviť konzistentnú sadu pravidiel, zjavne ide o nejaký druh merania a môžeme teda pristúpiť k zaujímavejšej otázke, a to o aký druh merania ide*“ (Stevens, 1946, s. 680). V porovnaní s klasickou teóriou merania je záver o možnosti merania v psychológii a teda aj o validite úplne opačný. Otázka kvantitatívnej štruktúry psychických atribútov stratila akúkoľvek relevanciu (viď Michell, 1999). Keďže každý psychologický test s konzistentnou procedúrou skórovania (napr. spočítavanie správnych odpovedí) je meraním, otázkou už je len, akú škálu charakter meracieho nástroja produkuje (Haig, Borsboom, 2008). Čo sa týka validity, z klasickej teórie testov (CTT), najčastejšie aplikovaného psychometrického modelu operujúceho na základoch Stevensovej teórie merania, definične vyplýva, že $E(X) \equiv T$, t.j. že každý test (resp. očakávaná hodnota skóre) je validný na meranie svojho pravého skóre (viď Lord, Novick, 1968)⁵⁾. Zároveň konceptuálne základy a ani syntax CTT neumožňuje formálne rozlíšenie latentného skóre konštruktu od pravého skóre. Právě skóre totiž navyše okrem cieľového konštruktu a iných relevantných faktorov väčšinou zahŕňa chyby merania, ktoré sú systematického charakteru.

Istým riešením tohto problému je teória zovšeobecniteľnosti (Cronbach, Nageswari, Gleser, 1963), ktorej CTT je špeciálnym prípadom. V rámci štúdie zovšeobecniteľnosti je tak zodpovedajúcim výskumným plánom a následnou analýzou variancie pozorovaného skóre možné explicitne modelovať systematické zdroje chybovej variancie (tzv. fazety) ako aj ich interakcie. CTT ako aj teória zovšeobecniteľnosti avšak nedokážu poskytnúť odhad úrovne latentnej premennej u konkrétneho subjektu, pri-

⁵⁾ Chybové skóre (E) je z definície náhodné a teda nezávislé na pravom skóre (T), preto platí, že $X = T + E$. Keďže predpokladaná náhodnosť E vyplývajúca z teórie chýb vedie v rade meraní k očakávanej hodnote $E = 0$, očakávané pozorované skóre X je rovné T. Z toho vyplýva, že výrok „X (DSB) meria T (WM)“ platí definične.

čom ich cieľom je iba zhodnotenie kvality pozorovaného skóre. V rámci CTT sa táto kvalita týka výlučne reliability. Viacero metód jej odhadu avšak vedie z pohľadu syntaxe CTT k navzájom nekompatibilným konceptom ako interná konzistencia, stabilita v čase, či zhoda medzi pozorovateľmi. Teória zovšeobecniteľnosti naopak dokáže formálne integrovať tieto rôzne zdroje chyby do jediného koeficientu. Ani jedna z týchto alternatív modelu pravého skóre však nedokáže interne reprezentovať vzťah medzi latentným skóre konštruktu a pozorovaným skóre. Právě skóre – či už ho vnímame z hľadiska CTT ako očakávanú hodnotu zo série na sebe nezávislých opakovaných meraní, alebo ako tzv. „skóre univerza“, kde pozorované skóre položiek a aspekty testovania sú de facto výberovým súborom z univerza možných pozorovaní – totiž nie je totožné s latentným skóre konštruktu. Dôvodom je, že testový výkon spravidla nie je realizáciou iba jedinej premennej, čo je základným predpokladom CTT (viď Lord, Novick, 1968). Keďže ani korelácia medzi pravým a pozorovaným skóre nemôže byť mierou validity (keďže takto je už v CTT definovaná reliability), tak priradenie významu pravého skóre, resp. validita, musí byť reprezentovaná externe, t.j. ako kovariancia s inou premennou. V rámci teórie zovšeobecniteľnosti sú možnosti zaradenia iných relevantných faziet do výskumného plánu často značne obmedzené, čo redukuje validitu na zovšeobecniteľnosť testového skóre naprieč rôznymi spôsobmi testovania (rôzne položky, hodnotitelia, čas testovania). To však nereflektuje otázku validity merania, t.j. aký konštrukt test meria. Psychometrické modely vychádzajúce z konceptu pravého skóre, ktoré operujú v rámci reprezentacionalistickej koncepcie merania, tak nedokážu napraviť neschopnosť reprezentacionizmu definovať validitu merania inak ako axiomatically.

Niekde uprostred predošlých dvoch teórií merania (klasickej teórie a Steversonho reprezentacionalizmu) stojí tretia, teória latentných premenných. Na jednej strane je táto teória merania postavená na reprezentacionalistických princípoch v tom zmysle, že jej syntax umožňuje formulovanie latentných premenných, ktoré sú iba elementmi matematického modelu reprezentujúceho sadu atribútov, resp. položiek (tzv. formatívne meranie). Vo všeobecnosti avšak táto teória priamo implikuje kauzálnu interpretáciu vzťahu atribút – indikátor (viď Bollen, 1989; Borsboom et al., 2003) a to najmä vo svetle syntakticky jasne definovanej jednosmernej vzťahovej väzby – od latentnej premennej k indikátorom (tzv. reflektívne meranie). Táto teória vníma atribút ako latentnú premennú, ktorá je nedokonale meraná sadou indikátorov. Teória latentných premenných predpokladá existenciu chýb merania a explicitne modeluje očakávané hodnoty pozorovaných premenných. Dokáže sa tak lepšie vyrovnáť s empirickými dátami, kde premenné sú zaťažené chybou merania, pričom výkon v indikátoroch, resp. odpovede na položky, sú funkciou latentnej premennej. Konštrukty tu teda nie sú determinované pozorovanými premennými, ale naopak, latentná premenná syntakticky zastupujúca daný konštrukt produkuje varianciu v jej indikátoroch, vždy avšak iba na pravdepodobnostnej úrovni. Pomocou psychometrických latentných modelov, t.j. teórie odpovedi na položku (IRT)⁶⁾ a faktorových modelov je možné empiricky testovať, či je variancia v indikátoroch determinovaná unidimenzionálne, t.j. ak indikátory merajú iba danú latentnú premennú a stávajú sa po parcializácii jej vplyvu navzájom nezávislými, je ich s istou rezervou (štatistická unidimenzionalita \neq kauzálna unidimenzionalita, viď napr. van der Maas et al., 2006) možné považovať za meranie jedného a toho istého atribútu.

⁶⁾ V rámci IRT je navyše možné odhadnúť presnosť merania v závislosti od úrovne latentnej premennej, pričom funkcia popisujúca vzťah medzi pravdepodobnosťou vyriešenia položky a latentným znakom (resp. obťažnosťou položky) môže nadobúdať rôzne formy.

Validita je následne vecou skúmania identity danej latentnej premennej, pričom, na rozdiel od Stevensovej reprezentacionalistickej teórie merania, validita testu na meranie latentnej premennej (resp. validita na meranie pravého skóre v prípade Stevensovej teórie) nie je daná definične, ale na jej preukázanie je potrebné verifikovať unidimenzionalitu celkového skóre.

Z uvedeného je zrejmé, že pravdivosť výroku „DSB meria WM“ závisí na arbitrárnej voľbe teórie merania, pričom rôzne teórie implikujú diametrálne odlišný záver ohľadom možnosti validity a vyžadujú použitie rôznych postupov jej dokladovania.

4 VALIDITA: RELAČNÝ VS. KAUZÁLNY KONCEPT

Je zjavné, že inferencia vedúca od pozorovania (DSB) k atribútu (WM) vyžaduje istý druh vzťahovej väzby. Charakter tejto väzby je determinovaný primárne interpretáciou konceptu validity. Hoci koncepcie validity definujúce danú väzbu ako meranie implikujú jej kauzálny charakter, vnímanie validity ako kauzálneho alebo relačného konceptu je do značnej miery arbitrárnou voľbou (rovnako, ako voľba teórie merania).

Ak vnímame validitu ako relačný koncept, tak overenie výroku „DSB meria WM“ vyžaduje v prvom rade externú reprezentáciu konštruktú. Predstavme si, že nájdeme kritérium Y (zastupujúce WM) a odhadneme populačnú hodnotu korelácie skóre testu DSB a kritéria, povedzme $\rho_{DSB,Y} = .50$. Znamená týchto 25% zdieľanej variancie to, že kritérium Y a DSB merajú ten istý konštrukt? Ak áno, je známa identita tohto konštruktú? Alebo kritérium Y a DSB merajú rozdielne konštruktú, ktoré sú avšak obe ovplyvnené treťou, neznámou premennou? Korelácia je druhom neanalyzovanej asociácie a tak tento model neobsahuje dostatočné množstvo informácií na to, aby bolo možné rozlíšiť medzi týmito možnosťami. Korelácie iba prinajlepšom indikujú validitu, nedokážu však jednoznačne uchopiť význam konštruktú a ani hypotetické poznanie všetkých korelátov z vedeckého hľadiska nepredstavuje dostačujúci dôkaz validity merania.

Ak meranie vnímame v súlade s princípmi exaktných vied ako kauzálny koncept, taktiež validizácia vyžaduje dôkazy o kauzálnom vplyve atribútu na merací nástroj. Vnímanie validity ako kauzálneho konceptu zároveň ide ruka v ruke s pozíciou realizmu v otázke existencie psychologických atribútov, keďže len to, čo existuje, môže byť kauzálne efektívne. Primárnym, resp. ideálnym nástrojom validizácie kauzálnych výrokov v rámci aktuálne najširšie akceptovanej teórie kauzality (tzv. kontrafaktuálna teória kauzality, vid' Lewis, 2001) je pritom pravý experiment (vid' ale Pearl, 2009 pre možnosť vyvodzovania kauzality v rámci neexperimentálnych plánov). Ten je založený na náhodnom pridelení úrovni nezávislej premennej (atribútu), t.j. napríklad zostavení experimentálnej a kontrolnej skupiny a overení, či rôzne úrovne nezávislej premennej determinujú úroveň efektu (výstupu merania), resp. či determinujú pravdepodobnostnú distribúciu výstupu merania (pri probabilistickom vnímaní kauzality)⁷⁾.

Problémom pri kauzálnom vnímaní validity však je, že v psychológii je v prípade organizmického typu premenných spravídla problematické manipulovať nezávislými premennými a získať tak relevantné empirické dôkazy. Napr. na preukázanie platnosti výroku „DSB meria WM“ by bolo potrebné subjektom náhodne priradiť istú úroveň pracovnej pamäti a zistiť, či tá kovariuje so skóre validizovaného testu. Ak sa takto preukáže, že test je citlivý na variáciu v danom atribúte, je možné prehlásiť

⁷⁾ Potreba probabilistickej teórie kauzality sa zvykne interpretovať ako dôsledok nedokonalosti deterministických závislostí alebo eventuálne tým, že samotný kauzálny systém má probabilistickú povahu.

výrok „DSB meria WM“ za pravdivý (viď Borsboom et al., 2004). Ak teóriu istého konštruktú alebo atribútu predstavuje iba nomologická sieť (napr., že WM pozitívne koreluje s krátkodobou pamäťou, selektívnou pozornosťou a Gf, a nekoreluje s motorickou rýchlosťou a extroverziou), tak jej vedomosť nijakým spôsobom neindikuje možnosť ani spôsob manipulácie s úrovňou daného atribútu. V mnohých prípadoch avšak možnosť manipulácie atribútu existuje, no pre jej znalosť je potrebná teória *fungovania* daného atribútu (viď Embretson 1998; Leighton, Gierl, 2007).

Pre ilustráciu, v prípade WM je jednou z takýchto teórií Baddelleyov WM model (2000). V rámci tohto modelu je pracovná pamäť systémom troch subsidiárnych komponentov; tzv. fonologickej slučky, vizuálno-priestorového náčrtníka, epizodického zásobníka a centrálnej exekutívy, ktorá manažuje prvé tri subsidiárne systémy. Vhľad do fungovania atribútu WM následne umožňuje použitie jednej z experimentálnych techník, ako je napríklad tzv. dual-task, prostredníctvom ktorej je možné selektívne blokovat' jeden zo subsystémov WM a testovat', ako toto vyradenie pracovnej pamäti vplýva na výkon v istej úlohe⁸⁾. Blokovanie zodpovedajúceho subsystému WM v experimentálnej skupine asociované so zníženým výkonom v teste DSB (prirodzene, odpoveďová modalita musí byť rozdielna) tak predstavuje kauzálny dôkaz jeho validity pre meranie atribútu WM. V rámci kauzálneho poňatia validity, poznanie, čo test meria, teda vyžaduje nie znalosť korelácií, ale teóriu fungovania daného atribútu v kontexte požiadaviek danej úlohy (Borsboom, 2005).

Výhodou relačného poňatia validity je nenáročnosť jej dokladovania. Na druhej strane, validita ako kauzálny koncept prestáva byť neuchopiteľnou a legitímne umožňuje silnú interpretáciu testového skóre ako manifestácie meraného atribútu. Pri iných ako kognitívnych atribútoch avšak často nie je známa možnosť akejkoľvek experimentálnej intervencie a kauzálna interpretácia testového skóre býva postavená na empiricky netestovaných predpokladoch. V tom prípade často praktizované pripísanie kauzálneho významu faktorovému modelu nestačí a je potrebné aspoň testovanie invariance daného modelu merania u rôznych populácií, ktoré vykazujú rôzne úrovne relevantných kontrolných premenných (viď Markus, Borsboom, 2013). Pri následnej interpretácii je však nutné vnímať limitácie vyplývajúce z kváziexperimentálneho pôvodu takýchto dôkazov.

ZÁVER

Kam až teda siahajú korene absencie konsenzu v otázke validity? Analýza vývoja tohto konceptu ako aj porovnanie dvoch súčasných pozícií vnímania validity (relačnej a kauzálnej) naznačuje, že absencia konsenzu nespočíva len v terminologických aspektoch a jej korene siahajú až k samotnej filozofii vedy. Kvôli snahe vysporiadať sa s problémami dokladovania validity sa tento koncept pohyboval na rozhraní dvoch vedeckých orientácií, scientizmu, usilujúceho o vedeckosť a praktikalizmu, riešiacieho praktické problémy. Výsledkom toho je, že v súčasnosti neexistuje univerzálne akceptovaný konsenzus v chápaní validity ba ani v otázke, na čo sa má tento koncept vzťahovať.

Hoci potreba konceptu validity historicky vyplynula z nejasností v otázke merania (Michell, 2011), vývoj validity priniesol namiesto adresovania týchto problémov skôr ich opomínanie (aplikáciou axióm operacionalizmu) a zároveň rast definičného oboru

⁸⁾ Vykonať istých, kognitívne veľmi nenáročných (a s inými kognitívnymi funkciami neinterferujúcich) úloh ako napríklad opakovaním „bla bla“ alebo ťukaním prsta po stole do tvaru štvorca je možné interferovať s činnosťou fonologickej slučky spracovávajúcej verbálne stimuly, resp. vizuálno-priestorového náčrtníka.

validity, ktorá dnes v rámci majoritného teoretického konsenzu zahŕňa väčšinu relevantných aspektov merania (viď Kane, 2006; Messick, 1989; AERA, APA, NCME, 1999). Avšak nie sú to len široko konceptualizované interpretácie validity, ktoré trpia teoretickou nešpecifickosťou a nejednoznačnosťou v otázkach jej dokazovania. Aj tá najjednoduchšia a najtechnickejšia interpretácia konceptu validity nie je invariantná, keďže vyžaduje arbitrárnu voľbu vnímania meraného konštrukt (realizmus vs. konštruktivizmus), ako aj interpretáciu pojmu merania (klasická teória merania, reprezentacionalizmus, teória latentných premenných). Na tom, či vnímame na základe týchto volieb validitu ako relačný alebo kauzálny koncept, možno však v konečnom dôsledku z pragmatického hľadiska až tak nezáleží. Keďže dáta pochádzajúce z validizačných štúdií sú primárne len čísla a tie si nepamätajú odkiaľ pochádzajú (viď Lord, 1953), aplikáciou psychometricko-štatistických modelov sa rozdiely medzi relačným a kauzálnym poňatím validity vytrácajú. Psychometrické modely reflektované v podobe kovariančných štatistických modelov totiž nemusia byť viazané na žiadnu formálnu teóriu vysvetľujúcu kauzálny pôvod testového skóre. Takto je možné bez teórie meraného konštrukt prehlásiť, že skóre je jeho manifestáciou alebo mierou, a nevnímať nutnosť špecifických dôkazných postupov (intervencie na úrovni konštrukt alebo na úrovni testovej situácie, resp. stimulov) potrebných pre oprávnenosť takejto interpretácie. Tento fakt môže byť rovnako pozitívom (z uhla praktikalizmu) alebo negatívom (z uhla scientizmu). Tak, ako viaceré volieb v rámci vnímania validity, aj táto je arbitrárna.

LITERATÚRA

- AERA, APA, NCME (1999): Standards for educational and psychological testing. Washington DC, Author.
- Anastasi, A. (1950): The concept of validity in the interpretation of test scores. *Educational and Psychological Measurement* 10, 67-78.
- Baddeley, A. D. (2000): The episodic buffer: A new component in working memory? *Trends in Cognitive Sciences* 4, 417-423.
- Barrett, P. T. (2005): What if there were no psychometrics? Commentary on Robert McGrath's „Constructs, complexity, and measurement“ target article. *Journal of Personality Assessment* 85, 134-140.
- Bickhard, M. H. (2001): The tragedy of operationalism. *Theory & Psychology* 11, 35-44.
- Bollen, K. A. (1989): *Structural equations with latent variables*. New York, Wiley.
- Borsboom, D., Cramer, A. O. J., Kievit, R. A., Scholten, A. Z., Franic, S. (2009): The end of construct validity. In: Lissitz, R. W. (Ed.), *The concept of validity: Revisions, new directions, and applications*. Charlotte, NC, Information Age, 135-170.
- Borsboom, D., Mellenbergh, G. J., Van Heerden, J. (2003): The theoretical status of latent variables. *Psychological Review* 110, 203-219.
- Borsboom, D., Mellenbergh, G. J., van Heerden, J. (2004): The concept of validity. *Psychological Review* 111, 1061-1071.
- Bridgman, P. W. (1927): *The logic of modern physics*. New York, Macmillan.
- Bunge, M. (2008): *Causality and modern science*. New Brunswick, NJ, Transaction Publisher.
- Carnap, R. (1936): Testability and meaning I. *Philosophy of Science* 3, 419-471.
- Colliver, J. A., Conlee M. J., Verhulst, S. J. (2012): From test validity to construct validity...and back? *Medical Education* 46, 366-371.
- Coombs, C. H. (1964): *A theory of data*. New York, Wiley.
- Cronbach, L. J. (1988): Five perspectives on validation argument. In: Wainer, H., Braun, H. (Eds.), *Test Validity*. Hillsdale, NJ, Erlbaum, 3-17.
- Cronbach, L. J. (1949): *Essentials of psychological testing*. New York, Harper.
- Cronbach, L. J. (1980): Selection theory for a political world. *Public Personnel Management* 9, 37-50.
- Cronbach, L. J., Meehl, P. E. (1955): Construct validity in psychological tests. *Psychological Bulletin* 52, 281-302.
- Cronbach, L. J., Nageswari, R., Gleser, G. C. (1963): Theory of generalizability: A liberation of reliability theory. *The British Journal of Statistical Psychology* 16, 137-163.
- Cureton, E. E. (1951): Validity. In: Lindquist, E. F. (Ed.), *Educational measurement*. Washington, DC, ACE, 621-694.

- Embretson, S. E. (1998): A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods* 3, 380-396.
- Haig, B. D., Borsboom, D. (2008): On the conceptual foundations of measurement. *Measurement* 6, 1-6.
- Hempel, C. G. (1965): *Aspects of scientific explanation*. New York, Free Press.
- Hölder, O. (1901): Die Axiome der Quantität und die Lehre vom Mass, *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig. Mathematisch-Physische Klasse* 53, 1-46.
- Hood, S. B. (2009): Validity in psychological testing and scientific realism. *Theory and Psychology* 19, 451-473.
- Kane, M. T. (2001): Current concerns in validity theory. *Journal of Educational Measurement* 38, 319-342.
- Kane, M. T. (2006): Validation. In: Brennan, R. (Ed.), *Educational Measurement*. Washington DC, ACE, NCME, 17-64.
- Kelley, T. L. (1927): *Interpretation of educational measurements*. New York, Macmillan.
- Leighton, J. P., Gierl, M. J. (2007): Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educational Measurement* 26, 3-16.
- Lewis, D. (2001): *Counterfactuals*. Malden, Wiley-Blackwell.
- Loevinger, J. (1957): Objective tests as instruments of psychological theory. *Psychological Reports* 3, 635-694.
- Lord, F. M. (1953): On the statistical treatment of football numbers. *American Psychologist* 8, 750-751.
- Lord, F. M., Novick, M. R. (1968): *Statistical theories of mental test scores*. Reading, MA, Addison-Wesley.
- Markus, K., Borsboom, D. (2013): *Frontiers of validity theory: Measurement, causation, and meaning*. New York, Routledge.
- McGrath, R. E. (2005): Conceptual complexity and construct validity. *Journal of Personality Assessment* 85, 112-124.
- Messick, S. (1989): Validity. In: Linn, R. L. (Ed.), *Educational measurement*. Washington, DC, ACE, NCME, 13-103.
- Michell, J. (1999): *Measurement in psychology: A critical history of a methodological concept*. New York, Cambridge University Press.
- Michell, J. (2005): The logic of measurement: A realist overview. *Measurement* 38, 285-294.
- Michell, J. (2011): Qualitative research meets the ghost of Pythagoras. *Theory and Psychology* 21, 241-259.
- Newton, P. E., Shaw, S. D. (2013): Standards for talking and thinking about validity. *Psychological Methods* 18, 301-319.
- Pearl, J. (2009): *Causality: models, reasoning and inference* (2nd ed.). Cambridge, UK, Cambridge University Press.
- Rogers, T. B. (1995): *The psychological testing enterprise: An introduction*. Pacific Grove, CA, Brooks.
- Sijtsma, K. (2009): Correcting fallacies in validity, reliability, and classification. *International Journal of Testing* 9, 167-194.
- Stevens, S. S. (1946): On the theory of scales of measurement. *Science* 103, 677-680.
- Strauss, M. E., Smith, G. T. (2009): Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology* 5, 89-113.
- Trendler, G. (2009): Measurement theory, psychology and the revolution that cannot happen. *Theory and Psychology* 19, 579-599.
- Urbánek, T., Denglerová, D., Širůček, J. (2011): *Psychometrika*. Praha, Portál.
- van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., Raijmakers, M. E. J. (2006): A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychological Review* 113, 842-861.
- Zumbo, B. D. (2009): Validity as contextualized and pragmatic explanation, and its implications for validation practice. In: Lissitz, R. W. (Ed.), *The concept of validity: Revisions, new directions, and applications*, Charlotte, NC, Information Age, 65-82.

SÚHRN

Príspevok sa zaoberá analýzou konceptu validity a hľadáním dôvodov absencie konsenzu v jej definovaní a vnímaní. Najvýznamnejšie koncepcie validity sú konfrontované z diachronického hľadiska s dôrazom na konceptuálne a epistemologické otázky. Táto analýza naznačuje, že rozdiely medzi jednotlivými koncepciami validity nie sú terminologickými hrami, ale majú vecné dôsledky. Zároveň je demonštrované, že význam aj tej najužšej a najčastejšie uvádzanej definície (test je validný, ak meria to, čo má merať) závisí na niekoľkých voľbách, ktoré sú v zásade arbitrárne. Tie následne implikujú empirické postupy potrebné pre atribúciu prívlastku „validný“ a v konečnom dôsledku aj to, či by mala byť validita považovaná za relačný alebo kauzálny koncept.

Copyright of Ceskoslovenska Psychologie is the property of Institute of Psychology of the Academy of Sciences and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.