

Validita a reliabilita výskumných nástrojov: princípy a reálna prax

Peter Gavora

Univerzita Tomáše Bati ve Zlíně, Fakulta humanitních studií, Centrum výzkumu

Redakci zasláno 1. 4. 2013 / upravená verze obdržena 4. 6. 2013 / k uveřejnění přijato 24. 6. 2013

Abstrakt: Článok odpovedá na otázku, ako sa uplatňujú v praxi princípy zisťovania validity a reliability, ktoré udávajú renomované učebnice metodológie. Prvé dve časti príspevku načrtávajú teoretický rámec a opisujú koncept validity a reliability. Nasledujúce časti vysvetľujú proces analýzy a uvádzajú jej výsledky. Z časopisu *The Journal of Educational Research* bolo náhodným spôsobom vybratých 56 štúdií. Veľká väčšina (91 %) výskumných nástrojov v nich boli posudzovacie škály a testy, zvyšok boli dotazníky, pozorovacie schémy a interview. Na prekvapenie bola validita zdokumentovaná len u 26 % nástrojov, zvyšok nástrojov boli štandardizované testy, alebo sa autori spoliehali na zrejmú (face) validitu. U posudzovacích škál sa použila konštruktová validita zistená pomocou faktorovej analýzy. Obsahová validita a face validita sa použili pri testoch, dotazníkoch a interview. Kombinované zdroje validity (napr. konštruktová spolu so súbežnou alebo diskriminačnou) sa použili veľmi málo, čo považujeme za nedostatok vo validizačnom procese. Reliabilita bola uvedená pri 80 % nástrojoch. Najfrekvencovanejší spôsob bola Cronbachova alfa u posudzovacích škál. Zhoda medzi posudzovateľmi sa použila pri pozorovaní a testoch. Test-retest sa použil pri overovaní, či je merací nástroj na pretest-posttest stabilný v priebehu času. Koeficienty reliability väčšinou presahovali 0,80. Pri analýze štúdií sa potvrdilo, že pri posudzovaní adekvátnosti validizácie a reliabilizácie je potrebné silne prihliadať k špecifickým podmienkam konkrétneho výskumu.

Kľúčové slová: metaštúdia, validita, reliabilita, výskumný nástroj, výskumná štúdia

Validita a reliabilita sú základné pojmy psychometrie, teórie vyjadrujúcej princípy a postupy objektívneho merania premenných veličín súvisiacich s človekom. Psychometria má za sebou storočnú tradíciu, ktorú založil F. Galton. Jej teóriu postupne rozvinuli osobnosti zaoberajúce sa meraním inteligencie, ďalších schopností, ako aj osobnostných črt (A. Binet, L. L. Thurstone, K. Pearson, W. Stern, Ch. Spearman, R. B. Cattell, L. J. Cronbach a ďalší). Validita a reliabilita sa viazala najprv na testy, ktoré boli sprvoti najpoužívanejšími a najrozvinutejšími objektívnymi výskumnými nástrojmi, neskôr k nim pri-

budli i dotazníky, posudzovacie škály a štruktúrované pozorovanie. Výskumné nástroje využívajúce tieto metódy zberu dát sa použili v experimentoch, t.j. výskumnej metóde, ktorá sa najviac chce priblížiť modelu výskumu reprezentovanému prírodnými vedami.

Ak skúmame pohyb v metodológii za posledných 30–40 rokov, zistíme, že v kvantitatívnom výskume zďaleka nebol taký dynamický ako v kvalitatívnom výskume. V kvantitatívnom výskume môžeme spomenúť napríklad novú teóriu testových odpovedí (*item response theory*), Raschov model merania, zisťovanie veľkosti účinku (*effect size*), ktoré sa stalo základom pre metaanalýzy a v štatistike hierarchickú regresnú analýzu. S príchodom počítačov sa rozvinulo adaptívne testovanie a elektronické administrovanie dotazníkov a posudzovacích škál, kde papier a ceruzku nahradila klávesnica a obrazovka. Zmenil sa však nielen nosič, ale aj forma odpovedí a spôsob navigácie vo výskumnom nástroji, zrýchlilo a zautomatizovalo sa spracovanie dát. Veľké zmeny priniesli možnosti použitia videa pri pozorovaní (v ČR najmä Janík a jeho skupina – Janík & Miková, 2006, Najvar et. al., 2011). Nové impulzy pre komparatívny kvantitatívny výskum priniesli medzinárodné štúdie gramotností (čitateľskej, matematickej, prírodovednej). Istý posun nastal v chápaní validity a reliability výskumného nástroja, o ktorom píšeme v nasledujúcej časti.

Kvantitatívny výskum si udržal silné postavenie vo výskumných štúdiách na celom svete, aj keď sa objavil jeho nový súper – kvalitatívny výskum založený na úplne inej filozofii a z nej odvinutej metodológii. Kvalitatívna metodológia najprv predstavovala rivala, neskôr sa stala partnerom a napokon spojencom kvantitatívneho výskumu. Objavili sa výskumné projekty, v ktorých bola prvá fáza kvantitatívna, druhá kvalitatívna a vice versa, alebo obidve fázy prebiehali simultánne (pozri Hendl, 2005). Neskôr koexistencia týchto dvoch prístupov poskytla pôdu na rôzne formy a úrovne prelínania. Napríklad pri naratívnej metóde, ktorá je jednou z vlajkových lodí kvalitatívneho prístupu, sa objavil i jej kvantitatívny variant – kvantitatívna naratíva. Naopak, zo štruktúrovaného dotazníka, ktorý je typickým predstaviteľom kvantitatívneho prístupu, sa môžu konštruovať „kvalitatívne“ naratívy. Sú založené skôr na časových a procesných charakteristikách než na interpretatívnych a subjektívne významových pozíciách a slúžia na vytvorenie „generických životných príbehov“ (Elliot, 2012).

V situácii dynamického metodologického vývoja je dobré vrátiť sa k základným kameňom kvantitatívneho výskumu a skontrolovať stabilitu jeho štan-

dardných postupov uplatnených vo výskumnej praxi. V tomto príspevku to uskutočníme prostredníctvom jeho dvoch kľúčových konceptov – validity a reliability.

1 Koncept validity a reliability

Validita a reliabilita sú kľúčovými požiadavkami kladenými na výskumný nástroj v kvantitatívnom výskume. Spolu s ďalšími vlastnosťami, ako je objektivita, verifikovateľnosť a štatistická zovšeobeciteľnosť zistení vytvárajú piliere toho, čo nazývame dobrý kvantitatívny výskum.

Validita sa v minulosti definovala ako vlastnosť výskumného nástroja umožňujúca zisťovať to, čo výskumník mieni zisťovať. Ak chce výskumník zistiť spokojnosť učiteľov s profesiou, do výskumného nástroja vloží položky, ktoré zahrňujú rôzne aspekty práce učiteľa a dá učiteľom možnosť, aby tieto aspekty hodnotili na škále spokojnosť-nespokojnosť. Z obsahového hľadiska potom možno kvalifikovať tento výskumný nástroj ako validný na zisťovanie spokojnosti učiteľa s profesiou. Novšia koncepcia validity výskumného nástroja (*Standards...*, 1999)¹ posunula tento koncept ďalej a výskumník sa na základe zistení pýta, ako môže interpretovať odpovede konkrétnej skupiny učiteľov z hľadiska ich spokojnosti s profesiou a aké má dôkazy na doloženie svojej interpretácie. Úroveň validity určitého výskumného nástroja je teda daná interpretačnými možnosťami jeho zistení. Dôležitú úlohu pri nej hrá vyvodzovanie, ktoré výskumník uskutočňuje na základe údajov zistených výskumným nástrojom.

Výskumník sa môže rozhodnúť pre rôzne spôsoby („zdroje“ v terminológii *Standards...*, 1999) validity. Z hľadiska toho, ako daný výskumný nástroj odráža obsah určitej výchovno-vzdelávacej oblasti možno usudzovať o obsahovej validite. Z hľadiska toho, do akej miery výskumný nástroj predikuje budúci stav alebo zisťuje súvislosť s danou vlastnosťou meranou iným nástrojom, možno usudzovať o kritériálnej validite. Z hľadiska toho, aký psychický konštrukt výskumný nástroj sleduje, možno usudzovať o jeho konštruktivej validite.

Reliabilita vyjadruje súčasne presnosť a spoľahlivosť zistených údajov. Výskumník má možnosť zvoliť si rôzne spôsoby reliability podľa toho, aké

¹ Ide o štandardy pre pedagogické a psychologické meranie, ktoré vydávajú spoločne Americká asociácia pre pedagogický výskum, Americká psychologická asociácia a Národná rada pre meranie vo výchove, a ktoré sú v USA rešpektovanou normou pre tvorbu a používanie výskumných a diagnostických nástrojov.

závery chce vysloviť. Ak chce vysloviť záver o tom, aké stabilné sú údaje získané pomocou výskumného nástroja, potom si zvolí opakovanie administrácie s tými istými subjektmi po časovom odstupe (*test-retest reliability*). Ak chce vedieť, do akej miery možno jeho dáta zovšeobecňovať na podobné úlohy, použije alternatívne formy výskumného nástroja s tými istými subjektmi. Ak chce vedieť, do akej miery sú dáta zhodné, keď ich vyhodnocujú viacerí hodnotitelia, porovná výsledky týchto hodnotiteľov (*inter-scoring reliability*). Ak chce vedieť, do akej miery je jeho výskumný nástroj homogénny (t. j. či sa jeho položky dostatočne koncentrujú na cieľovú vlastnosť), vypočíta korelácie medzi každou položkou a nástrojom ako celkom. Aby bola reliabilita dostatočná, musia byť korelácie dosť vysoké (Salvia & Ysseldyke, 1998, s. 151–152).² Vypočíta sa tiež index vyjadrujúci vnútornú konzistenciu celého výskumného nástroja, obyčajne je to koeficient alfa³.

Informácie o úrovni validity a reliability pomáhajú výskumníkovi rozhodovať sa o tom, či môže byť s výskumným nástrojom spokojný, alebo či ho má zlepšovať, prípadne odmietnuť. Tieto údaje potrebuje výskumník na začiatku výskumu. Ale otázky validity a reliability ho musia zaujímať aj pri spracovaní a interpretovaní výskumných dát. Dáta totiž musí vzťahovať vzhľadom k vlastnostiam výskumného nástroja.

Obidve tieto vlastnosti výskumného nástroja sú relatívnymi mierami. Nemôžeme tvrdiť, že je výskumný nástroj validný alebo je nevalidný, ale aký je stupeň jeho validity. Tiež je zrejmé, že výskumný nástroj môže byť viac validný na jeden účel a menej validný na iný účel. To isté platí o reliabilite. Vždy ide o údaj o konkrétnom výskume s konkrétnymi výskumnými osobami v konkrétnych podmienkach administrácie, to znamená, že údaje získané tým istým výskumným nástrojom sa môžu v rôznych výskumoch líšiť. V dobre overenom výskumnom nástroji a pri podobných skúmaných osobách a pri podobných podmienkach administrácie by však rozdiely nemali byť veľké.

Okrem výrazov validita a reliabilita budeme v tomto príspevku používať aj výrazy validizácia a reliabilizácia, ktoré sa týkajú procesu smerujúcemu k získaniu informácií o týchto vlastnostiach výskumných nástrojov. Validita a reliabilita sú výslednými produktmi tohto procesu.

² Nie však extrémne vysoké (napr. nad 0,95), pretože to naznačuje, že položky sú takmer totožné. A to zase spochybňuje validitu nástroja.

³ Klasici metodológie výskumu Nunnally a Cronbach (cit. podľa Kline, 2000, s. 43) tvrdili, že je to najlepší index reliability vôbec. Keďže je jeho výpočet aj pohodlný, spájajú sa v ňom dve ideálne vlastnosti.

2 Ciele štúdie

Metodologické princípy zisťovania validity a reliability výskumného nástroja, ich predpoklady, základné vlastnosti a normy dostatočnej validity a reliability sú známe z renomovaných publikácií (napr. *Standards...*, 1999; Kline, 2000; Hopkins, 1998). Aká je však skutočná prax? Ako sa tieto princípy uplatňujú v reálnych výskumoch s konkrétnymi výskumnými nástrojmi?

Odpoveď na tieto otázky je možné získať prostredníctvom analýzy produktov výskumníkov, teda publikovaných výskumných štúdií a správ. V tomto príspevku analyzujeme časopisecké výskumné štúdie, ktorých výber je uvedený nižšie. Žáner tohto príspevku charakterizujeme ako metaštúdiu, t. j. štúdiu, ktorá súhrnným spôsobom podrobuje analýze a hodnoteniu väčší počet iných štúdií z vopred stanovených hľadísk.

Pri analýze sme sa zamerali na tieto otázky:

1. Aké druhy výskumných nástrojov sa v štúdiách použili a v akej proporcii?
2. Koľko výskumných nástrojov obsahujú jednotlivé štúdie?
3. Aké spôsoby validizácie a reliabilizácie sa použili?
4. Bola validizácia a reliabilizácia použitého výskumného nástroja dokladovaná v danej výskumnej štúdii takým spôsobom, aby sa mohlo usudzovať, že sa uskutočnila vhodne a korektne?
5. Bola validita a reliabilita použitého výskumného nástroja dostatočne vysoká vzhľadom k platným princípom?

Prvé dve otázky sa týkajú informácií o zastúpení výskumných nástrojov v jednotlivých štúdiách. Mapujú základný kontext, v rámci ktorého bude prebiehať ďalšia analýza. Majú ukázať, ktoré typy výskumných nástrojov prevládajú a, naopak, ktoré sú slabo zastúpené (a prečo). Tretia otázka smeruje ku kategorizácii jednotlivých spôsobov validity a reliability. Štvrtá a piata otázka smeruje k jadrú príspevku a je zameraná na zistenie, do akej miery je prax validity a reliability verná platným metodologickým princípom.

3 Výber štúdií na analýzu

Odpoveď na naše výskumné otázky možno nájsť v publikovaných výskumných správach alebo časopiseckých štúdiách. V tomto príspevku sme sa

z pragmatických dôvodov sústredili na časopisecké štúdie – sú dostupnejšie než výskumné správy.

Pri ich výbere je možné postupovať tak, že sa zvolí pedagogický časopis (alebo niekoľko časopisov) rôznej orientácie a tieto sa podrobia analýze z uvedeného hľadiska. Podmienkou však je, aby časopis publikoval dostatočné množstvo výskumných štúdií. Ak je týchto štúdií málo, analýza neposkytne dostatočný obraz o realite. Okrem toho výskumné štúdie musia obsahovať opis zisťovania validity a reliability použitých výskumných nástrojov, bez nich by totiž nebolo čo analyzovať. Kritérium dostatočného množstva výskumných štúdií plnia mnohé, najmä zahraničné časopisy. Po zvážení možností sme vybrali jeden z nich, ktorý mal aj ďalšie ukazovatele výborné.

Zvolili sme časopis *The Journal of Educational Research*, a to z týchto dôvodov:

1. Časopis vychádza dlhodobo (v r. 2013 ide o 106. ročník).
2. Rukopisy sú prísne recenzované.
3. Publikujú v ňom významní výskumníci.
4. Aj keď v ňom prevládajú severoamerickí autori, redakcia uverejnila aj príspevky z iných geografických oblastí (Európa, Austrália, Ázia).
5. Časopis vychádza v známom vydavateľstve Taylor and Francis.
6. Takmer všetky príspevky sú zamerané na empirický výskum, takže časopis poskytne dostatok materiálu na analýzu.

Pri výbere konkrétnych štúdií sme postupovali nasledovne. V ročníkoch 2005, 2007, 2009 a 2011 sme zoradili štúdie abecedne podľa priezviska prvého autora. V každom ročníku sme vybrali prvých 15 štúdií, ktoré sa stali predmetom analýzy. Ak štúdia nezodpovedala kritériu, vybrali sme v poradí ďalšiu štúdiu. Kritériom bola orientácia na kvantitatívny výskum a použitie výskumného nástroja. Z výberu boli preto vyradené štúdie, ktoré obsahovali demografické analýzy, metaanalýzy a, samozrejme, kvalitatívne štúdie. Zo štúdií so zmiešaným, kvantitatívno-kvalitatívnym dizajnom bola predmetom analýzy len kvantitatívna časť. Aby sme sa vyhli efektu opakovania pri výbere štúdií, autor mohol mať vo výbere len jednu štúdiu, v ktorej bol uvedený ako prvý autor.

Časopis je dostupný v databáze *ProQuest Central* (prvé tri zvolené ročníky) a na webovej stránke časopisu (ročník 2011). Z posledného ročníka bolo voľne dostupných, resp. zodpovedalo kritériu len 7 štúdií.⁴ V starších ročníkoch boli niektoré štúdie nečitateľne naskenované, tie nemohli byť analyzované (poradie výberu sa posunulo). Za poznámku stojí, že časopis len výnimočne uverejňuje štúdie, ktorých hlavným cieľom je vývoj nového výskumného nástroja a podrobná správa o jeho validizácii a reliabilizácii. Publikuje hlavne štúdie, v ktorých sa výskumný nástroj používa primárne na zber výskumných dát, pričom opis a interpretácia týchto dát tvorí jadro výskumnej štúdie.

Celkove súbor zahrňuje 52 výskumných štúdií, v ktorých sa použilo 174 výskumných nástrojov.

4 Zistenia

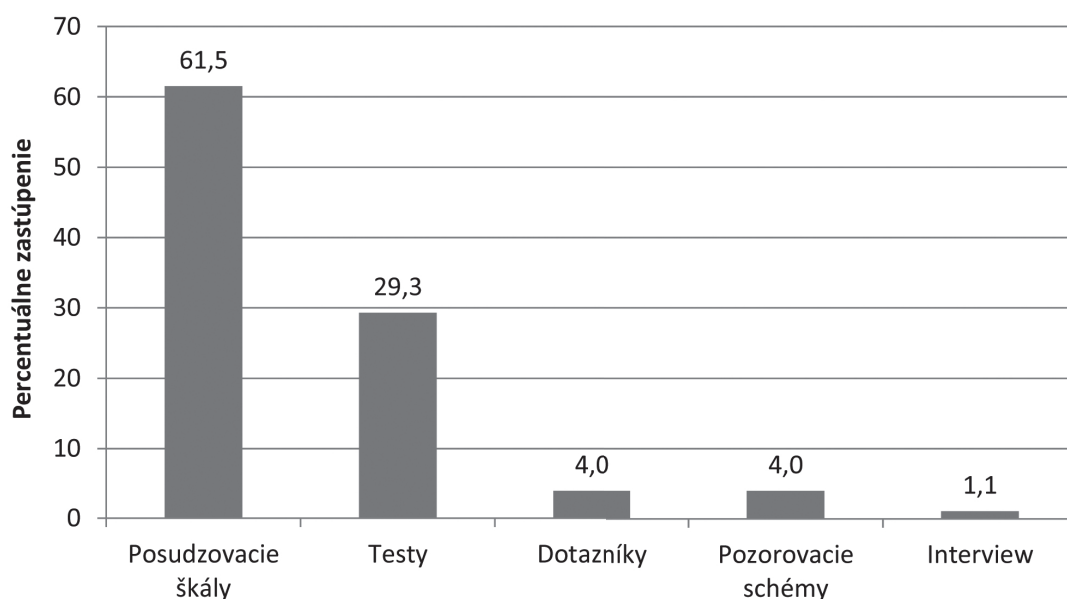
4.1 Druhy výskumných nástrojov

Zistenia o druhoch použitých výskumných nástrojov informujú čitateľov o metodologických trendoch, ktoré sa presadzujú v danom časopise. Ale druhy výskumných nástrojov v štúdiách časopisu vypovedajú nepriamo aj o tom, aké mali autori štúdií možnosti pri ich validizácii a reliabilizácii, resp. ktorým možnostiam dali prednosť. Niektoré výskumné nástroje sa totiž prednostne validizujú a reliabilizujú jedným spôsobom, iné druhým. Základné údaje o počte jednotlivých druhov výskumných nástrojov v analyzovaných štúdiách vyjadruje Obrázok 1.

Kvôli odstráneniu terminologickej nejednoznačnosti uvedieme najprv, čo rozumieme pod konkrétnym názvom výskumného nástroja. Posudzovacia škála znamená nástroj obsahujúci intervalové alebo poradové škály vo forme formulára. Umožňuje vyjadrovať skóre za celý nástroj, alebo za jeho zložky (dimenzie). Dotazník znamená nástroj obsahujúci uzavreté alebo otvorené otázky. Niekedy bývajú v dotazníkoch aj škálové položky, ale málokedy prevládajú. (Štruktúrované) interview je, podobne ako dotazník, formou získavania informácií od respondentov v podobe otázok, avšak ústnym spôsobom.

⁴ Články z ročníkov 2012 a 2013 neboli dostupné v databáze *ProQuest Central* alebo v inej databáze a neboli ani voľne dostupné na webovej stránke časopisu. Vydavateľ poskytuje prístup k štúdiám z týchto ročníkov za poplatok USD 28 za jednu štúdiu. Vzhľadom na to, že analýza týchto ročníkov by nepriniesla iný obraz o výskumných nástrojoch, ako bol ten z nižších ročníkov, považovali sme zakúpenie článkov z týchto ročníkov za nevhodné. Preto sme ich analýzu neuskutočnili.

U dotazníkov a interview sa vyjadruje frekvencia alebo percentuálne zastúpenie odpovedí. Výrazom test označujeme nástroj obsahujúci úlohy (položky) na meranie kognitívnych a psycho-motorických výkonov subjektov. Výsledok testu sa vyjadruje pomocou skóre. Pozorovacia schéma je systém kategórií, v rámci ktorých výskumník zaznamenáva frekvenciu, intenzitu, kvalitu alebo trvanie pozorovaných javov alebo procesov. Výsledok pozorovania sa vyjadruje numericky v podobe frekvencií, aritmetických priemerov, indexov atď.⁵



Obrázok 1. Druhy použitých výskumných nástrojov.

Ako vidno, vo výskumoch sa robustným spôsobom uplatnili dva druhy výskumných nástrojov – posudzovacie škály a testy. Tie spolu pokrývajú 90,8 % použitých výskumných nástrojov. V 52 štúdiách našej vzorky sa posudzovacie škály použili 107-krát, testy 51-krát. V štúdiách boli veľmi slabo zastúpené dotazníky, pozorovacie schémy a interview.

⁵ Terminológia súvisiaca s rôznymi druhmi výskumných nástrojov v angličtine je dosť komplikovaná. Slovenským výrazom dotazník sa obyčajne prekladajú tri anglické výskumné nástroje: questionnaire (dotazník), checklist, inventory (inventár), scale (škála). Zlé je aj to, že anglický výraz scale (v jednotnom čísle) znamená vlastne súbor škál, nie jeden výskumný nástroj založený na škálovaní. Niekedy sa vo význame súboru škál dokonca používa výraz test (napr. *Test rodinného prostredia* J. Senku, *Záujmový test VRIOT*, *Test diferenciacie záujmov DIT* a pod.). Sociometrický test (v terminológii J. L. Morena) je vlastne posudzovacia škála. Pri analýze výskumných štúdií sme sa usilovali zistiť typické črty výskumného nástroja a podľa nich ho kategorizovať podľa tu uvedených definícií.

Otázkou je, prečo výskumníci preferovali posudzovacie škály a testy a, naopak, málo používali ďalšie druhy výskumných nástrojov. Výskumník si volí výskumné nástroje na základe cieľa výskumu a výskumných otázok, na ktoré hľadá odpovede. Pritom využíva všetky možnosti, ktoré mu empirický prieskum dáva (subjekty výskumu a inštitúcie) a je zároveň nimi obmedzovaný. Posudzovacie škály predstavujú nesmierne široký potenciál možností na zisťovanie vlastností osôb, procesov a produktov. V posudzovacích škálach, ktoré boli v našom súbore štúdií, sme našli veľmi pestré zameranie. Tu je niekoľko príkladov: miera integrácie inovácií do vyučovania učiteľom, osobnostné črty učiteľa, úroveň motivácie žiakov, postoje žiakov k vyučovaciemu predmetu, prístup žiakov k učeniu sa, miera pracovnej spokojnosti učiteľov, metakognitívna uvedomenosť žiaka, vnímaná profesijná zdatnosť učiteľa, konštruovanie roly učiteľa, vzťah škola-rodina, školské prostredie, materiálne vybavenie školy.

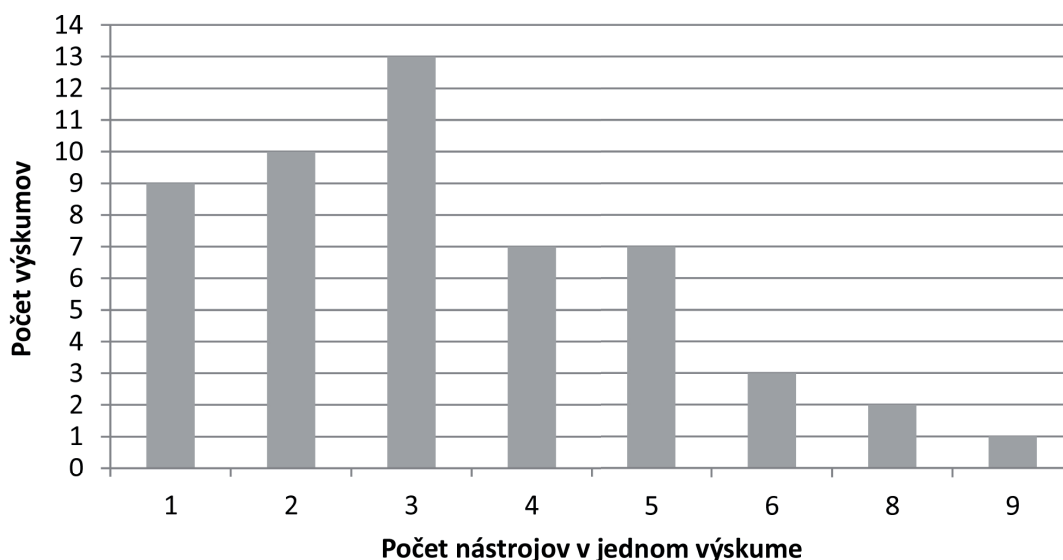
Okrem možností hodnotiť široké spektrum vlastností je výhodou posudzovacích škál relatívne ľahký spôsob zisťovania reliability. Vypočíta sa pomocou koeficientu alfa, ktorý obsahujú softvéry na štatistické spracovanie výskumných dát (SPSS, Statistica). Pri výpočte je zároveň možné dať príkaz na vypočítanie korelácie každej položky s nástrojom ako celkom (príp. s danou dimenziou). Ak je korelácia nízka, výskumník má možnosť pri príprave nástroja položku preformulovať alebo ju eliminovať, aby získal nástroj s vyššou konzistenciou, teda vyššiu reliabilitu. To je pomerne ľahká procedúra, ktorá – okrem možnosti skúmania širokého spektra vlastností – silne favorizuje tento výskumný nástroj. Ako ukážeme ďalej, zistenie reliability posudzovacích škál pomocou koeficientu alfa bolo najčastejším spôsobom reliabilizácie v našom súbore štúdií.

Testy sa v našom súbore štúdií vyskytli v počte 51, teda v priemere jeden test na jednu štúdiu. Pokiaľ ide o zameranie, boli to testy vedomostí a zručností z viacerých vyučovacích predmetov, testy gramotnosti na rôznej úrovni (pomenovanie písmen, dekódovanie, pomenovanie obrázka a pod.), testy divergentného myslenia, riešenia problémových úloh, predalgebraického myslenia, biligviálny test a pod. Mnohé použité testy boli štandardizované a bežne sa v USA používajú, napr. *SAT*, *KeyMath Revised Normative Update*, *Indiana Statewide Test for Educational Progress*, *North Carolina End of Course Algebra* a pod. Školstvo Spojených štátov, odkiaľ pochádzalo najviac výskumných štúdií, ktoré sme analyzovali, je výrazne výkonnostne orientované a testovanie je „denným chlebíčkom“ žiakov. To vysvetľuje silné zastúpenie testov v našom súbore štúdií.

Hoci sa dotazníky relatívne ľahko konštruujú (obyčajne ľahšie ako testy), nie sú zrejmé u výskumníkov publikujúcich v tomto v časopise populárne preto, lebo majú menšie štatistické možnosti. Nástroje na pozorovanie sa zvyčajne konštruujú ťažšie a zber dát je náročný na čas, podobne ako pri interview. Pravdepodobne sa preto použili menej často.

4.2 Počet výskumných nástrojov v jednej štúdií

Počet výskumných nástrojov v jednej štúdií vypovedá dostatočne silne o tom, aký široký bol výskumný záber. Pravdu povediac, zistenia o počte výskumných nástrojov v jednej štúdií boli pre nás väčším prekvapením, ako to, čo bolo hlavným predmetom analýzy, t. j. použité spôsoby a úroveň validity a reliability výskumných nástrojov.



Obrázok 2. Počet výskumných nástrojov v jednom výskume.

Obrázok 2 ukazuje, že vo veľkej väčšine štúdií sa použilo viac výskumných nástrojov ako jeden. V priemere to bolo 3,3 výskumných nástrojov na jeden výskum. Len deväť štúdií z 52 malo jediný výskumný nástroj; na druhej strane boli v našej vzorke výskumy, ktoré mali 4 až 9 výskumných nástrojov.

Počet a druh výskumných nástrojov závisí od cieľa výskumu a ten ukazuje, aké je široké (príp. hlboké) empirické pole, na ktoré sa výskum zameria. Z analýzy vyplýva, že časopis preferuje výskumy, ktoré danú problematiku skúmajú komplexne, berú do úvahy viaceré premenné a hľadajú medzi nimi

súvislosti. Na ilustráciu uvedieme výskum, ktorý je dosť typický pre takéto zameranie.

C. Kohova et al. (2009) skúmali vplyv skupinového učenia sa na motiváciu a učenie sa žiakov. Zisťovali názory žiakov o tom, či skupinová práca uspokojila ich potreby, ovplyvnila ich motiváciu a vytvárala dosť príležitosti na rozvoj zručností pre život. Výskumu sa zúčastnilo 588 žiakov stredných škôl a ich učitelia v Singapure. Výskumníci použili päť výskumných nástrojov adaptovaných z pôvodných anglických originálov:

- Posudzovacia škála motivácie (17 položiek).
- Posudzovacia škála uspokojovania potrieb (12 položiek).
- Posudzovacia škála skúseností s bádateľskými úlohami (17 položiek).
- Posudzovacia škála metakognície (7 položiek).
- Posudzovacia škála percepcie výsledkov učenia sa (15 položiek).

Okrem toho v kvalitatívnej časti výskumu autori použili pološtruktúrované interview s učiteľmi.

Na doplnenie analýzy počtu výskumných nástrojov použitých v jednom výskume uvedieme, že veľká väčšina štúdií mala viac ako jedného autora; nebolo neobvyklé nájsť štúdiu, ktorá mala štyroch až šiestich autorov. To naznačuje možnú spojitosť medzi počtom použitých výskumných nástrojov a počtom autorov. Výskum mal často teamové zameranie z dôvodu koncentrácie síl výskumníkov a efektívneho využitia prostriedkov.

4.3 *Originálne a prevzaté výskumné nástroje*

Výskumné nástroje, ktoré sa použili v našom súbore štúdií, boli buď vlastné (ich konštruktérmi boli autori štúdií), alebo boli prevzaté (vytvorili ich iní odborníci). Mohli to byť výskumné nástroje použité v iných výskumoch, alebo to boli štandardizované nástroje poskytované rôznymi vydavateľmi. Originálne výskumné nástroje boli zastúpené 47,1 percentami, prevzaté nástroje tvorili zvyšok (52,9 %). To, že prevzaté výskumné nástroje tvorili taký vysoký podiel, ukazuje, že výskumníci si mohli vyberať z bohatej ponuky, ktorú mali k dispozícii. Možnosť prevzatia originálneho výskumného nástroja v rodnom jazyku je veľkou výhodou, pretože uľahčuje prácu výskumníkov a umožňuje aj porovnávanie výsledkov rôznych výskumov realizovaných týmto nástrojom.

Prevzaté výskumné nástroje boli použité bez zmeny, alebo boli modifikované. Pri modifikácii výskumníci použili veľmi rôznorodé stratégie: zme-

nili znenie niektorých položiek nástroja, skrátili nástroj (vypustili niektoré položky), zlúčili dve dimenzie nástroja alebo prevzali len jednu dimenziu. Stávalo sa, že niekoľko skrátených nástrojov zlúčili do nového nástroja, pričom si prispôbali podľa potrieb aj ďalšie vlastnosti, napr. rozsah a znenie škály v posudzovacích škálach. Táto pestrosť zásahov a odvaha manipulovať s originálnymi nástrojmi nás tiež prekvapila, bola však obyčajne podložená dôkazmi o dostatočných psychometrických vlastnostiach modifikovaných nástrojov.

Ako ukážku výskumu založenom na množstve adaptovaných výskumných nástrojov uvedieme štúdiu H. J. Banga (2011). Autor skúmal, ako ovplyvňujú úroveň domáceho učenia sa žiaka činitele viažuce sa k samotnému žiakovi, jeho rodine a škole. Výskumu sa zúčastnilo 192 žiakov strednej školy z imigrantských rodín v USA. Autor použil sedem výskumných nástrojov, ktoré boli postupne žiakom zadávané:

- Posudzovacia škála angažovania sa do učenia adaptovaná z výskumu Suarez-Orozca a Suarez-Orozcovej (2001) – 10 položiek. Respondenti vyjadrovali frekvenciu rôznych činností, ktoré sú potrebné na úspech v škole. Príklad: „V škole dávam pozor.“
- Škála kognitívneho angažovania, ktorá bola zložená z položiek vybratých z dvoch nástrojov – 6 položiek. Položky zisťovali mieru intelektovej záťaže pri učení sa a záujem o učenie sa. Príklad: „Mám rád nové učebné výzvy.“
- Štýl domáceho učenia sa predstavoval skrátenú formu nástroja z výskumu Honga a Milgrama (2000) – 4 položky. Zisťoval preferencie žiaka pri učení sa. Príklad: „Mám rád také domáce zadania, pri ktorých môžem bádať.“
- Preferencia učebného prostredia pri učení sa doma predstavovala skrátenú formu nástroja z výskumu Honga a Milgrama (2000) – 4 položky. Položky zisťovali, v akom prostredí sa žiak rád učí. Príklad: „Rád si robím úlohy za písacím stolom.“
- Škála rodinných konfliktov bola adaptovaná z výskumu Prinza, Fostera, Kenta a O’Learyho (1979) – 11 položiek. Zisťovala konfliktné situácie medzi rodičom a dieťaťom ako napr. konflikty kvôli zlým známkam, množstvu domácich povinností dieťaťa a pod.
- Škála podporovania domáceho učenia sa žiaka rodičom bola prevzatá z výskumu Hoover-Dempseyovej et al. (2001) – 5 položiek. Príklad: „Ako často sa ti rodičia venujú pri domácom učení sa?“
- Škála násilia v škole bola prevzatá z dvoch výskumov – 7 položiek. Žiaci sa vyjadrovali, či musia robiť opatrenia, aby sa vyhli nebezpečným incidentom. Príklad: „Musím si zvoliť náhradnú cestu domov.“

Pre úplnosť dodávame, že okrem týchto siedmich nástrojov boli použité ďalšie tri, a to originály autora.

4.4 Dokladovanie validity

Zásady metodológie výskumu si vyžadujú, aby každý výskumný nástroj – to znamená aj prevzatý – mal zistenú validitu a reliabilitu v danom výskume. Je to preto, lebo je potrebné overiť, či výskumný nástroj obstoí v nových výskumných podmienkach (subjekty, prostredie, špecifiká administrácie, čas, ktorý uplynul od vzniku pôvodného nástroja). Výskumník musí brať do úvahy tieto okolnosti pri interpretácii svojich zistení. Validita a reliabilita sa nemusia zisťovať u štandardizovaných výskumných nástrojov, ktoré boli dostatočne vyskúšané a sú založené na reprezentatívnej výskumnej vzorke subjektov, s ktorou môže výskumník porovnať svoj výskumný súbor a na základe toho interpretovať svoje dáta. Autori štandardizovaných výskumných nástrojov by však mali v časovom odstupe overovať, či sú pôvodné normy ešte stále platné.

To, ako autori štúdie dokladujú postup pri validizácii svojho výskumného nástroja a ako opisujú výsledok tejto validizácie, je kľúčovým údajom na posúdenie toho, či bola táto validizácia uskutočnená vhodne a korektne.

Musíme začať zistením, či v štúdií autori vôbec uviedli údaje o validite svojich výskumných nástrojov. Vzhľadom na to, že sme tento časopis charakterizovali ako kladúci vysoké nároky na jeho autorov, môže sa zdať zvláštno, že v niektorých štúdiách tieto údaje neboli čitateľom poskytnuté. Ako uvidíme ďalej, situácia je zložitejšia.

Pokiaľ ide o validitu, 25,8% štúdií malo uvedený údaj, alebo to boli štandardizované nástroje, ktoré sú validné (boli validizované predtým). Zvyšok, to znamená veľká väčšina výskumných nástrojov údaj o validite nemala doložený. To je nepríjemné zistenie, ktoré však môžeme rozptýliť podrobnejšou analýzou. Našu pozornosť musíme diferencovať podľa skúmanej vlastnosti. Ak ide o vlastnosť alebo činnosť, ktorá je úplne zjavná, potom je zjavná aj validita daného nástroja. Test pravopisu vo výskume Madrida et al. (2007) je vlastne diktát, pri ktorom žiaci píšú vety, ktoré im diktujú. O jeho validite nemožno pochybovať⁶ a môžeme si ju overiť pohľadom (*face validity*). Test má zrejme aj dostatočnú reliabilitu. Podobne to môžeme povedať o rôznych testoch gramotnosti vo výskume Edmonsovej et al. (2009), kde sa použili subtesty ako pomenovanie písmen, obrázkov, aliterácia a rýmovanie.

⁶ Situáciu trochu zjednodušíme. Ak dáme žiakom diktát, v ktorom je veľa neznámych slov, potom to už nie je len test pravopisu, ale aj slovnej zásoby, príp. test odhadu významov, jazykového citu a pod. Tu už ide o inú vlastnosť, teda aj o inú validitu.

Inou kategóriou sú výskumné nástroje, ktoré boli adaptované (väčšinou skracované). Výskumníci validitu neskúmali, pretože dôverovali zisteniam konštruktérov o validite ich nástrojov. Vo veľkej väčšine prípadov však zisťovali ich reliabilitu, pretože tá závisí okrem iného aj od dĺžky výskumného nástroja. Ak výskumný nástroj je „katalóg“ (*checklist*) zisťujúci prítomnosť alebo neprítomnosť určitých objektov, napr. materiálneho vybavenia školy, validita je vysoká a netreba ju dokladovať. Niektoré výskumné nástroje nemali doloženú validitu, ale štúdiá obsahovali plné znenie výskumného nástroja, takže čitateľ si mohol utvoriť vlastnú mienku o ňom a iní výskumníci ho mohli použiť a overiť jeho psychometrické vlastnosti.

Spôsob validizácie závisí do značnej miery od druhu výskumného nástroja. V prípade posudzovacích škál bola v analyzovaných štúdiách vykazovaná zvyčajne faktorová analýza, čo je správny postup. Pomocou nej sa zisťuje konštruktívna validita, t. j. do akej miery sa položky tohto nástroja kryjú s daným konštruktom, ktorý leží v jeho základe a akú má výskumný nástroj faktorovú štruktúru (z akých dimenzií sa skladá) a či táto štruktúra zodpovedá teórii viažucej sa k danému konštruktu. Väčšinou sa použila exploračná faktorová analýza, v menšom počte prípadov konfirmačná faktorová analýza a ešte v menšom počte prípadov obidve. Faktorová analýza však predstavuje až poslednú fázu validizácie výskumného nástroja. Tejto fáze predchádza tvorba koncepcie výskumného nástroja, zostavovanie jeho položiek, ich obsahové overovanie, postupné doladovanie, zistenie, či zneniu položiek respondenti správne porozumeli atď. Táto fáza je oveľa dlhšia ako výpočet a interpretácia faktorovej analýzy, pretože zahŕňa aj pilotné overovanie v teréne, a to si obyčajne vyžaduje niekoľko cyklov.

Analyzované štúdie vo veľkej väčšine proces tvorby výskumného nástroja neopísali a vo výskumnej časti hneď prešli k údajom získaným z faktorovej analýzy (počet položiek, hranica faktorovej záťaže položiek, počet a označenie faktorov, celková vysvetlená variácia). Výnimkou z toho trendu je napríklad štúdiá Handelsmana et al. (2005), ktorí podrobne opísali proces generovania položiek nástroja. Použili však len jeden výskumný nástroj, takže mali dost textového priestoru – opis generovania zaplnil jednu tlačенú stranu. Ak by však autori použili napríklad 6 výskumných nástrojov, ako to bolo vo výskume Shihu (2009), potom ich textový priestor na opis validizácie by bol dost obmedzený. Shih použil konfirmačnú faktorovú analýzu a pri každom nástroji uviedol osem indexov, každý v rozsahu troch-štyroch riadkov. Dodajme, že išlo o prevzaté výskumné nástroje.

Napriek tomu sa domnievame, že obmedzenie rozsahom textu, ktorý mohli mať autori štúdií k dispozícii, nemusí byť jedinou príčinou, prečo proces tvorby a validizácie nástroja neopísali podrobnejšie. V rozvinutej metodologickej kultúre sa fáza tvorby, prípravy a overovania výskumného nástroja pred faktorovou analýzou považuje za samozrejmosť. Preto sa autori sústredili hlavne na opis poslednej validizačnej fázy, teda na faktorovú analýzu a prípravnú fázu neopísali.

Aby čitateľ mohol posúdiť, či bola faktorová analýza urobená dobre, je potrebné, aby štúdia poskytla podrobný opis postupu a jeho výsledkov. To bolo v štúdiách dosť typické. Dobrým príkladom je opäť štúdia Handelsmana et al. (2005).

Cieľom výskumu bolo zistiť mieru aktivity (angažovanosti, *engagement*) študentov na nižšom stupni vysokoškolského štúdia v USA. Na tento účel vyvinuli autori vlastnú posudzovaciu škálu, ktorá obsahovala 27 položiek. Dáta od 266 študentov podrobili exploračnej faktorovej analýze. Bola použitá metóda hlavnej osi a rotácia varimax, ktorá ponúkla štyri až sedemfaktorové riešenie. Na základe sutinového testu sa autori rozhodli pre 4 faktory, pri ktorých celková vysvetlená variácia bola 43 %. Autori uverejnili plné znenie položiek posudzovacej škály a faktorové záťaže v jednotlivých faktoroch (všetky nad 0,40). Okrem toho určili diskriminačnú validitu výskumného nástroja vzájomným korelovaním jeho dimenzií a konvergentnú validitu korelovaním skóre výskumného nástroja s prospechom študentov (tri hodnotenia študentov v priebehu semestra).

Takáto charakteristika je dostatočná na to, aby si čitateľ utvoril názor o spôsobe konštruktivej validizácie. Tu hovoríme len o požiadavkách pre časopisecské štúdie. Vo výskumných správach, v ktorých rozsah nie je obmedzený, by sa mali uviesť ešte podrobnejšie údaje. Podobné opisy validizácie poskytnú mnohé štúdie v našej vzorke.

Konštruktívna validita sa použila u 22 výskumných nástrojoch, z toho 7 bolo originálnych, 14 bolo prevzatých, ale validizovaných autormi štúdie na vlastnej vzorke. V jednej štúdii autori prevzali nástroj, ale uspokojili sa s údajom o validite zo vzorky konštruktérov výskumného nástroja. To je samozrejme jednoduchšie, ale menej korektné riešenie.

Pokiaľ ide o testy, k dispozícii je len 5 údajov o validite – u jedného testu sa použila face validita, u štyroch obsahová validita. Obsah položiek navrhovaného testu posúdili odborníci, pričom sa uviedla miera ich zhody v tom, ako položky testu pokrývajú štruktúru a obsah testovanej oblasti.

Duatepe-Paksu a Ubuz (2009) z Turecka použili vo svojom výskume dva vedomostné testy z geometrie. Ich obsahovú validitu posúdil učiteľ vyučovacieho predmetu a doktorand. Kritériom bolo pokrytie učiva testovými úlohami, kognitívna úroveň úloh a ich jazyková obťažnosť. Údaje o miere zhody posudzovateľov a o tom, ako sa riešila prípadná nezhoda však autori neuviedli. Spomenuli len spôsob validizácie.

V 46 testoch sa neuviedol nijaký údaj o validite, ale ako sme už spomenuli, mnohé z nich boli štandardizované alebo uznávané (*established*), iné testovali oblasti, kde face validita bola jasná.

Pokiaľ ide o dotazník, ten sa použil v našej vzorke výskumných štúdií sedemkrát a len v jednom prípade sa určila validita (obsahová), a to konsenzom učiteľov a žiakov. Pri výskumných nástrojoch založených na pozorovaní a interview nebola validita udaná.

Na záver časti o dokladovaní validity výskumných nástrojov v analyzovanom časopise kategorizujeme použité možnosti. Ak zoradíme jednotlivé spôsoby validizácie podľa dôkazovej sily, vychádza toto poradie. (V zátvorke sú počty výskumných nástrojov, pri ktorých sa daná validita aplikovala; pripomíname, že celkový počet nástrojov bol 174.)

1. Bola udaná konštruktová, obsahová alebo kritériálna validita (40). Pri dvoch výskumných nástrojoch autori uplatnili po dva zdroje validity.
2. Išlo o štandardizovaný nástroj, ktorý mal overenú validitu alebo to bol uznávaný (*established*) nástroj (5).
3. Face validita určená autormi štúdie (1).
4. Výskumný nástroj bol priložený k štúdii v plnom znení. Validita však nebola dokladovaná (10).

Prvý spôsob má najväčšiu dôkazovú silu. Validita sa určila pre konkrétny výskumný nástroj, konkrétnu vzorku subjektov a pre dané výskumné podmienky. Štandardizovaný výskumný nástroj je výborná alternatíva, ale dali sme ho na druhé miesto, pretože funguje dobre len vtedy, keď sa výskumná vzorka a podmienky výskumu nelíšia od tých, ktoré boli pri štandardizácii. Uznávaný výskumný nástroj je taký, ktorý bol mnohokrát použitý vo viacerých výskumoch, v ktorých sa vždy potvrdila jeho validita. S istou mierou neistoty jeho validita môže platiť aj v inom výskume, ktorý bol zrealizovaný za podobných podmienok aj keď sa pri ňom výpočty neuskutočnili. Uznávaný výskumný nástroj však na rozdiel od štandardizovaného nemá k dispozícii

normy, čiže jeho možnosti sú obmedzenejšie. Face validita je založená na kvalifikovanom hodnotení expertov; je v poriadku, keď nejde o ťažko identifikovateľnú vlastnosť, t. j. keď sa od posudzovateľov nevyžaduje vysoký stupeň vyvodzovania, ako je to napr. pri teste vedomostí z určitého učiva geografie. Pri zložitejších vlastnostiach, ako je napríklad vnímaná zdatnosť (*self-efficacy*) človeka pre určitú činnosť budú rozdiely medzi odhadom expertov väčšie. Ak je výskumný nástroj priložený k štúdiu, je to lepšie ako keď nie je dokladovaná nijaká validita. Čitateľ môže jeho validitu preskúmať. Existuje tiež šanca, že sa použije v iných výskumoch, kde sa jeho validita overí.

4.5 Dokladovanie reliability

Výskumný nástroj musí byť dostatočne reliabilný, aby s ním mohol výskumník pracovať, a preto je reliabilita výskumného nástroja rovnako dôležitá vlastnosť ako validita. Údaj o reliabilite bol uvedený u 79,6 % výskumných nástrojov.

Najviac bolo zastúpené vyjadrenie vnútornej homogenity výskumného nástroja pomocou koeficientu alfa – asi preto, lebo sa výhodne používa pri posudzovacích škálach, čo bol najčastejší typ výskumného nástroja. Koeficient alfa bol uvedený pri 91 nástrojoch. Z nich 31,7 % boli originálne nástroje autorov štúdií, 58,3 % tvorili prevzaté nástroje, ktoré však autori štúdií reliabilizovali na vlastnej vzorke respondentov, a len v 10 % prípadov sa autori štúdií uspokojili s reliabilitou, ktorú uviedli konštruktéri nástroja pri vlastnej vzorke respondentov. Tretí spôsob je najmenej vhodný. Samozrejme, ešte horší prípad je neuvedenie nijakého údaju o reliabilite (16 nástrojov). Za komentár stojí najmä vysoký počet výskumných nástrojov, ktoré boli prevzaté a reliabilizované v nových podmienkach. To je veľmi dobrá stratégia, pretože, ako sme uviedli, v nových podmienkach sa nástroj môže správať odlišne, a preto jeho psychometrické vlastnosti je potrebné vždy overiť. Koeficient alfa sa v analyzovaných štúdiách pohyboval väčšinou nad 0,80, teda dostatočne vysoko.

Na druhom mieste pokiaľ ide o frekvenciu spôsobu zisťovania reliability bola zhoda medzi posudzovateľmi (*inter-rater*), a to pri 21 výskumných nástrojoch. Použili sa jednak pri metóde pozorovania, jednak pri testoch. V druhom prípade išlo o úlohy, v ktorých sa skórovali odpovede na viacstupňovej škále. Pri takýchto typoch testových úloh je vždy vhodné si overiť, či sa tá istá odpoveď vyhodnocuje rovnako. Zhoda medzi posudzovateľmi (po zacvičení) bola veľmi vysoká – nad 0,90. Zisťovanie reliability opakovaným meraním po

časovom odstupe (*test-retest*) sa uplatnilo najmä vo výskumoch, kde subjekty prešli intervenciou (typicky vo vyučovacom experimente) a zisťovalo sa, či merací nástroj na pretest-posttest zostal dostatočne reliabilný aj na konci intervencie. Použité nástroje boli testy, posudzovacie škály, dotazníky alebo pozorovacie schémy (16 nástrojov). Reliabilita sa pohybovala od 0,68 do 0,95, väčšinou však nad 0,80.

Metodologické učebnice väčšinou uvádzajú minimálnu požiadavku 0,80 na reliabilitu výskumného nástroja. Koeficient 0,70 už naznačuje, že nástroj má isté problémy (Kline, 2000, s. 13–15). Pri údají 0,60 je potrebné k výskumnému nástroju pristupovať opatrne. Pri štandardizovaných testoch, ktoré sú dostatočne overené, je namieste požiadavka 0,90 a vyššia (Hopkins, 1998).

Toto sú učebnicové odporúčania. Výskum sa však uskutočňuje v konkrétnych podmienkach, ktoré ovplyvňujú reliabilitu, a preto údaj o reliabilite je potrebné interpretovať vždy z hľadiska týchto podmienok. Nástroje mávajú rôznu reliabilitu vzhľadom na ľudské vlastnosti, ktoré merajú. Test vedomostí z daného okruhu učiva by mal mať spravidla vyššiu reliabilitu ako posudzovacia škála záujmov a postojov a tá vyššiu ako test tvorivosti. Vplyvným faktorom je tiež obťažnosť položiek nástroja. Preto je veľmi dôležité, aby výskumník svoj nástroj opísal dost' podrobne, aby si čitateľ o ňom utvoril plastický obraz. Dobrou praxou je pripojiť k štúdiu plné znenie výskumného nástroja (bolo to v osemnástich štúdiách) alebo aspoň ukážky položiek (veľmi často). Niekde bol opísaný celý vývoj výskumného nástroja, i keď sa redakcia vyhýba uverejňovať takéto štúdie.

Ďalším faktorom, ktorý ovplyvňuje reliabilitu, je veľkosť a charakter výskumnej vzorky. Je známe, že u homogénnej vzorky býva reliabilita nižšia ako u nehomogénnej vzorky a vice versa. Test študijných schopností má u žiakov základnej školy vyššiu reliabilitu ako u žiakov gymnázia, pretože vzorka žiakov gymnázia je z hľadiska študijných schopností homogénnejšia ako u žiakov základnej školy. Pohľad na rozptyl skóre tiež pomôže interpretovať, prečo je reliabilita nízka. V každej z analyzovaných štúdií bola výskumná vzorka opísaná veľmi podrobne, a to nielen z hľadiska lokality, veku a genderu subjektov, ale aj ich etnickej príslušnosti. Ak to bolo potrebné z hľadiska zamerania štúdie, uviedli sa aj ďalšie ukazovatele, napr. poberanie školskej stravy zadarmo alebo za zníženú cenu, ako ukazovateľ sociálno-ekonomického statusu rodiny, percento detí so špeciálnymi potrebami, postihnutím, materinským jazykom iným ako angličtina a pod.

Počet položiek v nástroji je ďalším determinantom reliability. Dlhšie nástroje majú spravidla vyššiu reliabilitu ako kratšie nástroje. Dôležité je tiež rozlišovať, kedy autori uvádzajú reliabilitu za celý nástroj a kedy za jednotlivé dimenzie, pretože tie majú nárok na nižšiu reliabilitu. Dĺžka však nie je jediná vlastnosť nástroja, ktorá ovplyvňuje reliabilitu. Tá tiež závisí od počtu dimenzií, veľkosti ich vzájomnej korelácie a priemeru interkorelácií položiek v celom nástroji.

Seitsingerová (2005) použila vo svojom výskume nástroj nazvaný Classroom Instructional Practice Scale, ktorý zisťoval, v akej miere učitelia v USA pokrývajú predpísané učivo a aká je štruktúra ich vyučovacích postupov a stratégií. Išlo o sebahodnotiaci, prevzatý nástroj. Mal 16 dimenzií (napr. vyučovanie v malých skupinách, rozvoj kritického myslenia, komunitné učivo, vyučovanie občianskych a sociálnych zručností, učenie o zdraví, využívanie literárnych zdrojov a pod.). Celková reliabilita bola veľmi vysoká (0,95) zrejme aj preto, lebo nástroj bol veľmi dlhý, obsahoval 86 položiek. Reliabilita jednotlivých dimenzií bola v rozmedzí 0,24 a 0,79 (nie je udané, koľko položiek mali dimenzie).

4.6 Porovnanie vyjadrovania validity a reliability

Základným zistením je, že reliabilita bola vyjadrovaná oveľa častejšie ako validita, pričom pomer je približne 3 : 1. Prevalu vyjadrovania reliability môžeme vysvetliť jednoduchším spôsobom jej zisťovania najmä pri posudzovacích škálach, ktoré tvorili najfrekvencovanejší výskumný nástroj. Niektorí autori zisťovali reliabilitu výskumného nástroja viacerými metódami, pričom väčšinou použili výpočet vnútornej konzistencie a stability. Na druhej strane pri zisťovaní validity výskumných nástrojov sa viaceré spôsoby použili málokedy – konštruktová spolu so súbežnou len dvakrát, konštruktová spolu s diskriminačnou trikrát. Tento nízky počet tiež dokumentuje situáciu validity v porovnaní s reliabilitou výskumných nástrojov.

5 Diskusia

Naším cieľom bolo posúdiť validitu a reliabilitu výskumných nástrojov, ktoré boli použité vo vybraných štúdiách daného časopisu. Z piatich otázok položených na začiatku výskumu, na ktoré sme hľadali odpovede, priniesli tri také informácie, ktoré neboli celkom predpokladané.

Pokiaľ ide o druhy výskumných nástrojov v skúmaných štúdiách, bolo prekvapením veľmi vysoké zastúpenie posudzovacích škál a testov, ktoré pokryli

takmer 91 % použitých výskumných nástrojov. Výskumník si volí výskumný nástroj podľa cieľa svojho výskumu, ale svoj výskumný cieľ si nastavuje aj podľa dostupných výskumných nástrojov. Nemôže si totiž stanoviť cieľ, na riešenie ktorého nemá inštrumentálne možnosti. Do hry pri voľbe výskumných nástrojov vstupuje aj pragmatický aspekt. Posudzovacie škály a testy majú výborné možnosti hromadnej administrácie, dobré štatistické možnosti a možnosti zisťovania reliability (v porovnaní s interview a pozorovacími schémami). To ich zrejme favorizovalo pri výbere do konkrétnych výskumov. Veľké množstvo hotových posudzovacích škál a testov, ktoré autori štúdií prevzali z iných výskumov zdôvodňuje ich frekventovaný výber. Navyše, ako sme ukázali, výskumníci pristupovali k prevzatým nástrojom veľmi flexibilne – skracovali ich, kombinovali, zlučovali viaceré nástroje tak, aby to vyhovovalo cieľom výskumu.

Nepredpokladaným zistením bol veľký počet výskumných nástrojov v jednom výskume, čo ukazuje, že autori riešili komplexné, široké otázky, na zodpovedanie ktorých nevyhnutne potrebovali viaceré výskumné nástroje. Použili viaceré premenné, medzi ktorými hľadali vzťahy a spojenia. Túto snahu môžeme vsadiť do širšieho koncepčného rámca kvantitatívneho výskumu. Disman (1993) vo svojej téze o informačnej redukcii vyjadruje myšlienku, že žiadny výskum nemá možnosť uchopiť realitu v plnosti, pretože je nesmierne zložitá. Preto ju redukuje na skúmateľný úsek, zložky alebo rovinu. Nasadenie veľkého množstva premenných a početných výskumných nástrojov v skúmaných štúdiách je snahou zmierniť uvedenú redukcii informácií a postupnými krokmi sa priblížiť k plnšiemu opisu reality.

Tretia výskumná otázka sa týkala spôsobu validizácie a reliabilizácie výskumných nástrojov. Zistenia zodpovedali cieľom výskumu a možnostiam výskumných nástrojov. Najfrekventovanejším spôsobom validity bola konštruktová validita. Tu sa na chvíľu zastavíme. Ako je známe, tvorba výskumného nástroja (pre zjednodušenie budeme mať na mysli len posudzovacie škály) prebieha tak, že výskumník na základe svojej interpretácie príslušného konštruktú zostaví položky výskumného nástroja a podrobí ich validizácii pomocou faktorovej analýzy (pred ňou ešte uskutoční položkovú analýzu). Tu je potrebné pripustiť, že faktorová analýza by mala s istou pravdepodobnosťou potvrdiť autorovu interpretáciu, veď dané položky nástroja vznikli na jej základe (a nevhodné boli eliminované v procese položkovej a faktorovej analýzy). To ukazuje, že proces konštruktivej validizácie je akoby uzavretý do seba a sám seba potvrdzuje. Na to, aby autor zistil, či konštrukt dobre interpre-

toval pri navrhovaní položiek svojho výskumného nástroja a či boli položky dobre sformulované, by sa mal presvedčiť ešte zistením súbežnej a diskriminačnej validity. V prvom prípade ide o zistenie, či jeho výskumný nástroj dobre koreluje s inými nástrojmi, ktoré merajú totožný konštrukt. V druhom prípade ide o zistenie, či jeho výskumný nástroj nízko alebo záporne koreluje s výskumným nástrojom, ktorý meria úplne odlišný konštrukt. Týmto postupom by sa dokazovací kruh otvoril a konštruktová validita by sa mohla považovať za dokázanú. Ako sme uviedli v časti 4.6, súbežná a diskriminačná validita sa však zisťovala len v niekoľkých prípadoch, čo považujeme za slabinu validizačného procesu v skúmaných štúdiách.

Doklad o validizácii a reliabilizácii použitého výskumného nástroja (štvrtá otázka) je dôležitou informáciou vedúcou k zisteniu, či sa uskutočnila vhodne a korektne. Čitateľ časopisu by mal mať dostatok relevantných informácií o tom, ako validizácia a reliabilizácia prebehla a aký mala výsledok, aby mohol úroveň výskumu hodnotiť a interpretovať. S prekvapením sme však zistili, že takmer tri štvrtiny výskumných nástrojov nemalo uvedený údaj o validite, a tým ani o spôsobe validizácie. V texte sme vysvetlili, že tento na prvý pohľad hrubý nedostatok možno zmierniť, keď sa na výskumné nástroje pozrieme podrobnejšie. Mnohé z nich mali zrejmú (face) validitu, a to často formálne (expertne) potvrdenú. Osemnásť výskumných nástrojov (zo 174) bolo doložených in extenso; ak tak nebolo a išlo o nový nástroj, autori obyčajne uviedli aspoň príklady položiek. To je dobrý krok umožňujúci čitateľovi posúdiť validizáciu daného výskumného nástroja. Reliabilita bola dokumentovaná častejšie ako validita, čo možno pripísať okrem iného pohodlnej možnosti jej zisťovania u posudzovacích škál a testov, najfrekvencovanejších výskumných nástrojov v našej vzorke štúdií.

Pokiaľ ide o piatu otázku, či bola validita a reliabilita použitých výskumných nástrojov dostatočne vysoká, odpoveď je pozitívna. Veľká väčšina výskumných nástrojov mala koeficient reliability 0,80 a vyšší; jednotlivé dimenzie výskumných nástrojov však mali niekedy reliabilitu nižšiu, a to najmä kvôli menšiemu počtu položiek. Vysoká reliabilita výskumných nástrojov je dobrou správou, pretože ona je aj nevyhnutnou podmienkou ich dostatočnej validity (Kline, 2000, s. 15).

Pri celkovom hodnotení psychometrickej stránky výskumných nástrojov v našom výbere štúdií môžeme konštatovať veľmi dobrú úroveň. Ak sa vyskytli odchýlky od požadovaných princíпов, tieto boli prirodzené a zdôvodniteľné. Ukázala sa tiež stabilita postupov validizácie a reliabilizácie v čase.

Analyzované štúdie pokryli obdobie šiestich rokov; počas ktorých sme nezistili zmeny trendu v žiadnej zo skúmaných otázok tohto výskumu.

Zostáva ešte vyjadriť sa k motivácii tohto príspevku vyjadreného na jeho začiatku – či v podmienkach dynamického rozvoja metodológií zostali klasické postupy vyjadrovania validity a reliability výskumných nástrojov stabilné. Pohľad do analyzovaných čísiel časopisu ukazuje, že piliere kvantitatívnych výskumov zostali neporušené. Validita a reliabilita zostávajú nevyhnutnou súčasťou dobrých kvantitatívnych výskumov, ale pri jej posudzovaní je potrebné prihliadať k špecifickým podmienkam a možnostiam konkrétneho výskumu.

V texte sme sa vyhli explicitnému porovnávaniu validity a reliability výskumných nástrojov v analyzovaných štúdiách so situáciou v našich pedagogických časopisoch – a ani v tomto závere to nemienime urobiť. Nielen preto, že k dispozícii nemáme príslušné údaje na komparáciu, ale aj preto, lebo sa obávame, že mechanické porovnávanie s výskumami, ktoré vznikli za iných podmienok skrýva v sebe mnoho nástrah. Tým však netvrdíme, že by sa takáto analýza nemohla uskutočniť. V prvej aproximácii by sa však mala dať prednosť analýze na „vnútroštátnej“ úrovni.

Literatúra

- Bang, H. J. (2011). Newcomer immigrant students' perspectives on what affects their homework experiences. *The Journal of Educational Research, 104*(6), 408–419.
- Disman, M. (1993). *Jak se vyrábí sociologická znalost*. Praha: Vydavatelství Karolinum.
- Duatepe-Paksu, A., & Ubuz, B. (2009). Effects of drama-based geometry instruction on student achievement, attitudes, and thinking levels. *The Journal of Educational Research, 102*(4) 272–286.
- Edmonds, E., O'Donoghue, C., Spano, S., & Algozzine, R. F. (2009). Learning when school is out. *Journal of Educational Research, 102*(3), 213–221.
- Elliot, J. (2012). *Using narrative in social research. Qualitative and quantitative approaches*. Los Angeles: Sage.
- Handelsman, M. N., Briggs, W. L., Sullivan, N., & Towler, A. (2005). A measure of college student course engagement. *Journal of Educational Research, 98*(3), 184–189.
- Hendl, J. (2005). *Kvalitativní výzkum. Základní metody a aplikace*. Praha: Portál.
- Hong, E., & Milgram, R. M. (2000). *Homework: Motivation and learning preference*. Westport, CT: Bergin & Garvey.
- Hoover-Dempsey, K. V., Battiato, A. C., Walker, J. M., Reed, R. P., De-Long, J. M., & Jones, K. P. (2001). Parental involvement in homework. *Educational Psychologist, 36*(3), 195–209.
- Hopkins, K. D. (1998). *Educational and psychological measurement and evaluation*, 8th edition. Boston: Allyn and Bacon.

- Janík, T., & Miková, M. (2006). *Videostudie: výzkum výuky založený na analýze videozáznamu*. Brno: Paido.
- Kline, P. (2000). *Handbook of psychological testing*. 2nd edition. London: Routledge.
- Koh, C. K., Wang, J., Tan, O. S., Liu, W.C., & Ee, J. (2009). Bridging the gaps between students' perceptions of group project work and their teachers' expectations. *Journal of Educational Research*, 102(5), 334–347.
- Madrid, L. S., Canas, M., & Ortega-Medina, M. (2007). Effects of team competition versus team cooperation in classwide peer tutoring. *Journal of Educational Research*, 100(3), 155–160.
- Najvar, P., Najvarová, V., Janík, T., & Šebestová, S. (2011). *Videostudie v pedagogickém výskumu*. Brno: Paido.
- Prinz, R. J., Foster, S. L., Kent, R. N., & O'Leary, K. D. (1979). Multivariate assessment of conflict in distressed and non-distressed mother-adolescent dyads. *Journal of Applied Behavior Analysis*, 12(4), 691–700.
- Salvia, J., & Ysseldyke, J. E. (1998). *Assessment*. 7th edition. Boston: Houghton Mifflin Company.
- Seitsinger, A. (2005). Service learning and standards-based instruction in middle schools. *Journal of Educational Research*, 98(1), 19–30.
- Shih, S. S. (2009). An examination of factors related to Taiwanese adolescents' reports of avoidance strategies. *Journal of Educational Research*, 102(4), 377–388.
- Standards for educational and psychological testing*. (1999). Washington: American Educational Research Association.
- Suarez-Orozco, C., & Suarez-Orozco, M. (2001). *Children of immigration*. Cambridge, MA: Harvard University Press.

Autor

Prof. PhDr. Peter Gavora, CSc., Univerzita Tomáše Bati ve Zlíně, Fakulta humanitních studií, Centrum výzkumu, Mostní 5139, 760 01 Zlín, e-mail: gavora.p@gmail.com

Validity and reliability of research instruments: Principles and actual practice

Abstract: The aim of the paper was to explore how principles of estimating validity and reliability of research instruments, as described in respected methodology textbooks, are satisfied. The initial parts of the paper delineate the theoretical framework and describe the concepts of validity and reliability. The following sections of the paper explain the process of analysis and its findings. The *Journal of Educational Research* was chosen as the research focus. A sample of 56 randomly selected articles from it has been inspected. The analysis revealed that a large majority (91 %) of research instruments used in these articles was scales and tests, the rest were questionnaires, observation schemes and interviews. Surprisingly, validity was calculated only with 26 of instruments; the rest of instruments were standardized tests or they were face-

validated. As far as scales are concerned, construct validity was documented by means of factor analyses. Content validity and face validity were used in tests, questionnaires and interviews. We consider the infrequent use of combination of two sources of validity (e.g., construct and concurrent or discriminant) to be a weak element in the validation processes in the sample of studies. Reliability was documented with 80 % of research instruments. The most frequent method of calculation was Cronbach's alpha. Inter-rater reliability was used in observations and tests; test-retest reliability was used to control the stability of the pretest-posttest measuring instrument. The size of reliability coefficients in most of studies exceeded 0.80. Throughout the analysis it was corroborated that when judging validity and reliability one has to critically consider the specific conditions of each research study before expressing an evaluation statement.

Keywords: metastudy, validity, reliability, research instrument, research study

Krejčí, J., & Leontiyeva Y. (Eds.). (2012). *Cesty k datům. Zdroje a management sociálněvědních dat v České republice*. Praha: SLON.

Prostředí pro vědeckou práci v sociálních vědách prochází překotným vývojem. Jedním z důvodů je také rostoucí objem dat, která jsou k dispozici pro sociální analýzy. Využívání nových možností i zpřístupňování vlastních databází bývá podmínkou pro plnohodnotné zapojení do života mezinárodní vědecké komunity a pro dosahování výsledků s uznávaným vědeckým impaktem. Efektivní práce s daty však vyžaduje orientaci v nepřehledném množství zdrojů, zvládnutí metod managementu dat a sekundární analýzy i orientaci v právních a formálních náležitostech sdílení dat mezi výzkumnými týmy. Autoři a autorky této knihy se právě proto zaměřují na problematiku vytváření a používání kvantitativních sociálních dat, věnují se metodám managementu dat a představují služby datových archivů. Zvláštní pozornost věnují problematice harmonizace dat pro účely časových a mezinárodních komparací. Velkou část knihy tvoří přehledové studie o dostupných datech včetně hlubších rozborů a analýz datových zdrojů k vybraným tématům. Kniha je primárně zaměřena na akademický sociální výzkum, své čtenáře si ale jistě najde i ve výzkumu pro potřeby státní administrativy, v soukromých výzkumných agenturách i mezi novináři a všemi, kteří pracují s výstupy sociálních výzkumů. Jaké jsou zdroje dat v sociálních vědách? Kde hledat a jak získat data z předchozích šetření a sociálních statistik? Jak úspěšně realizovat vlastní výzkumný projekt? Jakých mezinárodních šetření se zúčastnila Česká republika? Jsou data z mezinárodních šetření srovnatelná? Jak měřit vzdělání a sociální postavení tak, aby byly výsledky srovnatelné? Kde najít data o gramotnosti populace, politických postojích, názorech na náboženství, manželství a rolích v rodině nebo data o migraci? Jaké databáze vytváří Český statistický úřad a k čemu jsou využitelné?