



**Sociálne a politické analýzy**

2012, 6, 2, s. 1-15

<http://sapa.ff.upjs.sk>

ISSN 1337 5555

## Validita, reliabilita a chyby merania

Alojz Ritomský<sup>1</sup>

Fakulta psychológie, Paneurópska Vysoká škola v Bratislave

---

### Abstrakt

Reliabilitou merania zisťujeme, do akej miery je výsledok merania určitej vlastnosti zaťažený chybou. Reliabilitu odhadujeme opakovaným meraním (stabilita), alebo na základe posúdenia toho, či každá časť merania dáva zhodné výsledky (interná konzistentnosť). Kritérium pre meranie reliability má dva aspekty: 1. testovanie je spojené s očakávaním stability, čo znamená, že jednotlivé výsledky meraní stálych vlastností sa nebudú meniť v čase ich merania, 2. konzistencia – aj jednotlivé zložky testu sú spojené s očakávaním ich presnosti. Argumentácia v prospech validity je najčastejšie založená: na obsahu testu, na vnútornej štruktúre testu, na vzťahu k iným premenným, na postupoch reagovania na testové položky a na dôsledkoch testovania. Aby testovanie malo pre diagnostiku alebo výskum význam, musí byť reliabilné a mať validnú interpretáciu testových hodnôt. Hlboká znalosť tejto problematiky je potrebná nielen pri vyvíjaní nového testu (merania), či jeho adaptácie, ale už pri štúdiu existujúcej testovej príručky a samotnej aplikácie testu. Testová, či užívateľská príručka je autormi a vydavateľmi testu pripravená publikácia, ktorá poskytuje informácie o celi testu, jeho vhodnom použití, o správnej administrácii a skórovaní, o normách, interpretácii výsledkov a prípadových štúdiách. Účelom tohto textu je priblížiť čitateľovi stručne problematiku predovšetkým k štúdiu testovej príručky.

---

**Kľúčové slová:** akademická veda, praktická veda, teoretické metódy, empirické metódy, praktické metódy

---

<sup>1</sup> Doc. PhDr. Alojz Ritomský, PhD., Fakulta psychológie, Paneurópska Vysoká Škola, Tematinska 10, SK-851 05 Bratislava, Slovenska Republika, e-mail: [ritomsky@is.paneurouni.com](mailto:ritomsky@is.paneurouni.com)

## Úvod

Otázkam merania sa v sociálnych vedách, či už vo svete alebo doma, venuje dlhodo- bo pozornosť. Existuje rozsiahly zoznam literatúry, ktorá rozvoj tejto oblasti zachytáva. V slovenskom jazyku sú to, napríklad, Ako sociologicky analyzovať (1976) od A. Hirnera, Kompendium psychodiagnostických metód od V. Černého a T. Kollárika (1986) a Metodologické základy psychologického výskumu od autorky L. Maršálovej (1978). V novšom období k riešeniu týchto otázok prispeli M. Hrabovský (1990) J. Schenk a A. Hra- bovská (2010) a ďalší. V češtine k starším pozoruhodným príspevkom patria diela týchto autorov: práca P. Řičana Úvod do psychometrie (1977), N. F. Kerlinger (1972), P. Blahuš (1996), J. Kožený (1998), J. Ferjenčík (2000), E. Reiterová (2003), A. Melka (2005), nevelké rozsahom, ale pozoruhodné dielo J. Srnca (2010) a napokon dôležitá monografia trojice T. Urbánek, D. Denglerová a J. Širuček Jan (2011). V poľštine medzi najvýznamnejšie k danej téme môžeme zaradiť dielo S. Nowaka (1975) a prácu J. Brzezińskiego (2003). V angličtine z rozsiahleho zoznamu spomenieme len klasikov S. S. Stevensa (1946), L. L. Thurstona (1959) a novšie, veľmi obľúbené dielo J. Michella (1999). Azda najväčší význam pre súčas- ného čitateľa majú Štandardy: APA (1985) a AERA, APA, NCME (1999). Posledné dielo je už viac než desať rokov k dispozícii aj v češtine pod názvom Štandardy pro pedagogické a psychologické testování (2001).

Problematika, ktorá stojí pred nami je teda naozaj rozsiahla. Tu sa môžeme pokúsiť ukázať v stručnosti len tie najpodstatnejšie poznatky. Medzi najzávažnejšie oblasti kvality merania psycho-sociálnych javov predovšetkým patria **reliabilita, validita a objektivita**. Tieto tvoria obsah nasledujúceho textu. Uvedieme ich stručnú charakteristiku.

**Reliabilita.** Reliabilitou merania zisťujeme, do akej miery je výsledok merania určitej vlastnosti zafaržený chybou. Reliabilitu nemôžeme hodnotiť priamo, pretože nepoznáme pra- vé výsledky. Takže ich odhadujeme opakovaným meraním (stabilita), alebo na základe posú- denia toho, či každá časť merania dáva zhodné výsledky (interná konzistentnosť). Kritérium pre meranie reliability (presnosť merania) má dva aspekty: 1. testovanie je spojené s očaká- vaním stability, čo znamená, že jednotlivé výsledky meraní stálych vlastností sa nebudú meniť v čase ich merania (opakované meranie vlastnosti za porovnateľných podmienok by mali viesť k rovnakým výsledkom), 2. konzistencia (vnútorný súlad) – aj jednotlivé zložky testu sú spojené s očakávaním ich presnosti. Vzhľadom k tomu, že test musí niečo merať, a neprodukovať významné chyby, hovoríme, že reliabilita je hranicou validity. Test, ktorý pro- dukuje veľké chyby nič nemeria (nie je reliabilný), preto nemôže byť ani validný.

**Validita.** Pod validitou (platnosťou) sa tradične rozumie schopnosť merania diagnos- tikovať, predikovať a merať tie javy, ku ktorým bol skonštruovaný. Kládne sa dôraz na to, čo sa meria. Podľa Kerlingera (1972, s. 435) najvšeobecnejšia definícia validity je v stručnosti

obsiahnutá v otázke: „Meriame to, o čom sa domnievame, že to meriame?“. Ako si neskôr ukážeme, obsah pojmu validita v ostatných rokoch prešiel dôležitými inováciami.

**Objektivita.** Objektivita sa týka zásad hodnotenia výsledkov merania. Má byť vymedzená tak, aby každá testujúca osoba, ktorá interpretuje rovnaký testový materiál prišla k rovnakým záverom a interpretáciám. Objektivita vyžaduje, aby test obsahoval zásady posudzovateľov reakcií a odpovedí pri prideľovaní číselných výsledkov a ich transformácie na špecifické normy (v hodnotení jednotlivca ako člena určitej referenčnej skupiny). Objektivita je navrhnutá tak, aby sa minimalizoval vplyv testujúceho na výsledky testu.

Hlboká znalosť tejto problematiky je potrebná nielen pri vyvíjaní nového testu (merania), či jeho adaptácie, ale už **pri štúdiu existujúcej testovej príručky a samotnej aplikácie testu**. Testová, či užívateľská príručka je autormi a vydavateľmi testu pripravená publikácia, ktorá poskytuje informácie o celi testu, jeho vhodnom použití, o správnej administrácii a skórovaní, o normách, interpretácii výsledkov a prípadových štúdiách (Standardy, 2001).

Účelom tohto textu je priblížiť čitateľovi stručne problematiku predovšetkým v kontexte uvedeného druhého cieľa, teda k štúdiu testovej príručky.

## I. Reliabilita

### A) Teória reliability v rámci klasickej teórie testu

Klasická teória testov je najčastejšou bázou pre odhad mnohých psychometrických charakteristík. Predpokladá, že skóre merania je sumou dvoch hodnôt, pravdivého a chybového skóre:

$$X = T + E,$$

kde

X – skóre merania,

T – pravé (pravdivé) skóre a

E – chybové skóre (majú náhodný charakter).

Teória predpokladá, že chyby merania majú náhodný charakter. Ďalšími sú predpoklady o skóre a faktoroch, ktoré ich ovplyvňujú:

1. priemer chybových skóre je nulový,
2. korelácia medzi pravým a chybovým skóre je nulová a
3. korelácia medzi chybovými skóre dvoch paralelných meraní je nulová.

Význam psychologických meraní spočíva v tom, že sme pomocou nich schopní odhadovať rozdiely medzi ľuďmi. Rozdiely medzi skóre rôznych ľudí sú výrazom rozdielov v ich pravých skóre a rozdielov medzi chybami merania.

Reliabilita zobrazuje vzťah medzi testovými, pravými a chybovými skóre. Existujú viaceré prístupy k reliabilite. Ukážeme si štyri *teoretické prístupy k analýze reliability*:

1. **Reliabilita ako pomer rozptylu pravých skóre k rozptylu testových skóre.** Hodnoty tohto pomeru sú v rozpätí od nuly po jednotku. Vysoké hodnoty ukazujú dobrú psychometrickú vlastnosť, pretože pri vysokých hodnotách veľký podiel rozdielov medzi skóre merania môže byť pripísaný rozdielom pravých skóre. Ak je rozptyl pravých skóre rovný nule – čiže všetci majú rovnaké pravé skóre, potom je reliabilita nulová. Aj je rozptyl pravých skóre rovný rozptylu skóre merania, potom je reliabilita rovná jednej. Niet chyby merania (testovania). Avšak, v realite je vždy prítomná určitá chyba merania.
2. **Reliabilita ako rozdiel medzi jednotkou a pomerom chybového k testovému rozptylu.** Ak je podiel *chybového k testovému rozptylu* nízky reliabilita je vysoká. Ak je rozptyl chybových skóre rovný nule – čiže všetci majú skóre merania rovné pravému skóre, potom je reliabilita rovná jednej. Aj je rozptyl chybových skóre rovný rozptylu skóre merania, potom je reliabilita rovná nule. Niet rozptylu pravého skóre. Rozdiely vo výsledkoch merania medzi ľuďmi predstavujú len náhodné odchýlky, nie skutočné rozdiely.
3. **Reliabilita ako miera, s akou rozdiely v skóre merania sú späté s rozdielmi v pravých skóre.** Koeficient korelácie medzi dvomi premennými vyjadruje silu vzťahu (spätosti) medzi nimi. V tomto zmysle môžeme aj reliabilitu chápať ako mocninu koeficientu korelácie medzi testovými a pravými skóre. Ak je takto chápaná reliabilita rovná hodnote jedna, potom možno povedať, že rozdiely medzi testovými skóre ľudí sú ideálne späté s rozdielmi medzi ich pravými skóre. Reliabilita rovná nule hovorí o tom, že rozdiely medzi skóre merania ľudí nesúvisia s rozdielmi medzi ich pravými skóre. Takéto meranie je zjavne neúčinné.
4. **Reliabilita ako miera, s akou skóre merania nesúvisia s chybou merania** (s chybovými skóre). Ak je korelácia medzi testovým a chybovým skóre nulová, reliabilita je maximálna. So vzrastaním hodnôt koeficientu korelácie reliabilita klesá. Ak je korelácia medzi testovým a chybovým skóre jednotková, reliabilita je minimálna.

### Štandardná chyba merania

Koeficient reliability je užitočný ukazovateľ, ktorý umožňuje porovnať reliabilitu viacerých testov. Pomocou neho sa dá rozhodnúť, ktorý test je reliabilnejší. Koeficient reliability, však priamo nerieši otázku veľkosti chyby merania. Veľkosť chyby merania je užitočná v interpretácii presnosti testových skóre. Štandardná odchýlka chybového skóre (odmocnina rozptylu) sa tu nazýva štandardnou chybou merania a označuje sa symbolom SEM. Štandardná chyba merania je priemernou hodnotou chyby merania. Čím je štandardná odchýlka merania väčšia, tým väčší je priemerný rozdiel medzi testovými a pravými skóre a tým menej je test reliabilný. Ako vidieť aj zo vzorca,

$$SEM = \sigma_x \cdot \sqrt{(1 - r_{xx})}$$

Ako vidíme, SEM závisí od reliability.

## B) Empíria reliability testu

Ukážeme si tri spôsoby odhadu veľkosti koeficientov reliability.

### a) Paralelné testy

Ako môžeme vidieť, teória reliability pracuje s pojmami: testové, pravé a chybové skóre. V konkrétnom meraní, však, ani pravé, ani chybové skóre nepoznáme. Ako sa teda dajú aplikovať spomínané teoretické poznatky o reliabilite s empirickým poznaním, zistením veľkosti reliability v konkrétnom testovaní? Odpoveď na túto otázku v rámci klasickej teórie testov vedie cez (ďalšie) predpoklady. Predpokladá sa, že môžu byť skonštruované dva psychologické testy, pre ktoré platí: obidva testy (formy) merajú tú istú psychickú vlastnosť (tomuto predpokladu sa hovorí tau-ekvivalencia), v obidvoch testoch sú pravé skóre rovnaké ( $T_1 = T_2 = T$ ). Predpokladá sa tiež, že obidva testy majú rovnaké chybové rozptyly. Takéto testy sa nazývajú paralelné testy. Potom z pohľadu klasickej teórie testov platí, že korelácia medzi dvoma paralelnými testami

$$X_1 = T + E_1$$

$$X_2 = T + E_2$$

je rovná ich reliabilite.

Ak máme testové skóre obidvoch foriem testu na tých istých osobách, môžeme vypočítať koreláciu medzi nimi a ňou odhadnúť reliabilitu testu.

Predpoklad, že skóre paralelných testov vyjadrujú tú istú psychologickú vlastnosť v praxi je splnený do určitého stupňa. Problémom môže byť skutočnosť, že formy testov sa líšia obsahom. Inak povedaná, že formy testu merajú odlišné psychologické konštrukty. Ak sú splnené predpoklady pre paralelné formy testu, priemerné hodnoty a štandardné odchýlky obidvoch testov sú si blízke. Koreláciu medzi paralelnými formami budeme nazývať reliabilitou paralelných foriem ( $r_{xx} = r_{AB}$ ).

### ***b) Test-retestova metóda***

Testy, ktoré merajú pomerne stabilné psychologické vlastnosti (napríklad inteligenciu, extraverziu a pod.) sa môžu vyhnúť niektorým problémom, s ktorými v reálnych aplikáciách zápasia paralelné formy. Metóda založená na opakovanom meraní je podobná s metódou paralelných foriem, keďže sa opiera na tie isté predpoklady. Predpokladáme, že pravé skóre testovaných osôb sa od prvého k druhému meraniu nemenia. Druhý predpoklad je, že chybový rozptyl prvého testovania je rovný chybovému rozptylu druhého testovania. Ak sú tieto predpoklady splnené, potom korelácia medzi prvým a druhým testovaním je odhadom reliability ( $r_{xx} = r_{x1x2}$ ).

Čím nižšiu hodnotu má koeficient korelácie, tým je väčší efekt chyby merania a tým nižšia je reliabilita testu. Svoju úlohu tu hrá čas medzi prvým a druhým testovaním.

### ***c) Reliabilita testu ako jeho vnútorná konzistencia***

Čo sa myslí pod vnútornou konzistenciou testu? Základná myšlienka tohto prístupu je založená na tom, že rôzne časti testu – jednotlivé položky alebo zoskupenia položiek možno skúmať ako rôzne formy testu. Platí, že ak sú položky medzi sebou silno späté, je pravdepodobné, že testové skóre sú konzistentné s pravými skóre a test pokladáme za reliabilný. Je tiež dôležité koľko položiek tvorí test. Dĺžka testu vplyva na veľkosť reliability. Dlhší test je reliabilnejší ako kratší test.

Ukážeme si dva spôsoby odhadu vnútornej konzistencie: 1. metódu rozpoľovanie, 2. výpočet koeficientu alfa.

#### ***Metóda rozpoľovania (Split-half)***

Položky testu rozdelíme do dvoch rovnakých skupín. Vypočítajme testové skóre pre každú skupinu. Máme dva subtesty, o ktorých budeme predpokladať, že majú vlastnosti paralelných testov. Delením položiek sa snažíme získať dve rovnocenné časti. V praxi často položky rozdelíme na párne a nepárne. Iný spôsob je, že položky do skupín vyberieme náhodne. Ďalší možný postup spočíva v tom, že vytvárame dvojice položiek s približne rovna-

kými psychometrickými vlastnosťami a potom náhodne jednotlivé položky pridelujeme skupinám.

Vypočítame koeficient korelácie medzi časťami testu ( $r_{12}$ ). Keďže korelačný koeficient nemá počet dvojíc hodnôt rovný celkovému počtu položiek testu, jeho veľkosť sa transformuje podľa vzorca:

$$r_{xx} = 2 \cdot \frac{r_{12}}{1 + r_{12}}$$

a tým získame odhad reliability testu.

Existuje zovšeobecnená formula vyššie uvedeného vzorca, ktorá má tvar:

$$r_{xx_n} = \frac{nr_{xx}}{1 + (n-1)r_{xx}},$$

kde  $r_{xx}$  je reliabilita pôvodného testu,  $r_{xx_n}$  je reliabilita predĺženého testu (zväčšeného o určitý počet položiek) a číslo  $n$  je násobok, o ktorý sa test predlžuje.

Z predchádzajúcej rovnice plynie, že výraz pre výpočet dĺžky testu na základe reliability pôvodného testu a požadovanej reliability má tvar:

$$n = \frac{r_{xx_n} \cdot (1 - r_{xx})}{r_{xx} \cdot (1 - r_{xx_n})}.$$

Zo vzorca plynie vyššie uvedené tvrdenie, že pri predĺžení testu reliabilita stúpa.

### **Meranie vnútornej konzistencie testu koeficientom alfa**

**Koeficient alfa pre viacvariantné položky** (položky s ordinálnou škálou):

$$\text{alfa} = r_{xx} = \frac{k}{k-1} \cdot 1 - \frac{\sum s_i^2}{s_t^2},$$

kde

$s_t^2$  - rozptyl testového skóre,

$s_i^2$  - rozptyl  $i$ -tej položky a

$k$  - počet položiek.

Štandardizované alfa:

$$\text{standardné alfa} = r_{xx} = \frac{k\bar{r}}{(1 + (k-1)\bar{r})}$$

kde

$k$  – počet položiek,

$\bar{r}$  - priemer koeficientov korelácií medzi položkami.

**Koeficient konzistencie** K-R 20 (podľa Kudara a Richardsona) **pre dichotomické položky:**

$$KR_{20} = r_{xx} = \frac{k}{k-1} \cdot \frac{s_t^2 - \sum p \cdot q}{s_t^2}$$

kde

$q = p - 1$ ,  $p \cdot q = s_i^2$ ,  $s_t^2$  - rozptyl testových skóre.

### Odhady veľkosti štandardnej chyby merania

V teoretických poznámkach o reliabilite sme uviedli vzorec pre štandardnú chybu merania (testu). Jej veľkosť môžeme odhadnúť z empirických dát pomocou vzorca:

$$\text{sem} = SD \cdot \sqrt{(1 - r_{xx})},$$

kde

$s_x$  je štandardná odchýlka hrubého skóre testu a

$r_{xx}$  je koeficient reliability.

## II. Validita

Testujeme to, o čom sa domnievame, že testujeme? Alebo trochu inak, testujeme to, čo chceme testovať? Sú to neľahké otázky. Test pracovného výkonu by mal merať pracovný výkon, a nie niečo iné. Požiadavka, otázka je triviálna, ale skutočná odpoveď na ňu je neľahká. Reč je o validite testu.

Meranie intelektu pomocou úloh, ktoré vyžadujú premýšľanie sa zdá byť vhodnejší, než zisťovanie, ako dokážu testovaní prejsť bicyklom medzi dvoma bielymi čiarami (Good-



win, 2002). Na prvý pohľad je riešenie úloh o usudzovaní pre testovanie intelektu validnejšia, než jazda na bicykli. Mnohé testy v populárnych časopisoch sú založené len na takomto zdôvodňovaní. Vedecké chápanie validity je však omnoho zložitejšie a komplexnejšie. K súčasnému chápaniu validity zásadným spôsobom prispievajú Štandardy... z roku 1999 (Štandardy, 2001). Americká psychologická asociácia (APA) kodifikuje **novú odbornú terminológiu** pre túto problematiku, ktorá má zabezpečiť **výrazne hlbší pohľad** na pojem validity v súvislosti s psychologickým a pedagogickým testovaním .

Validita je miera, s ktorou zhromaždené empirické poznatky a teórie podporujú špecifickú interpretáciu testových skóre, pri doporučenom spôsobe použitia testu. Inak vyjadrené: ide o špecifickú interpretáciu testových skóre, vyplývajúcu z navrhnutého spôsobu použitia testu.

„Nie je to teda test sám osebe, čo vyžaduje validizáciu, ale interpretácia skóre získaných pri jeho administrácii“ Štandardy...s. 17

Validizácia je proces získavania vedeckej evidencie, ktorá podporuje navrhnutú interpretáciu testových skóre a ich vhodnosť vzhľadom k navrhovanému použitiu testu.

Reč je o interpretácii testových skóre získaných doporučeným spôsobom použitia testu. Miera s akou je táto interpretácia podporená určitými empirickými a teoretickými poznatkami. Argumenty v prospech validity predstavujú „výslovné vedecké zdôvodnenie miery, v ktorej zhromaždené evidencie a teórie podporujú navrhnutú interpretáciu testových skóre“ (s. 173)

Rozlišujeme dva spôsoby interpretácie testových skóre:

- Normatívna interpretácia - interpretácia testových skóre vzhľadom k normám - „interpretácia testových skóre opierajúca sa o porovnanie výkonu testovanej osoby s výkonom osôb vo vymedzenej referenčnej populácii“ (s. 178),
- Kriteriálna interpretácia testových skóre: „umožňuje interpretovať testové skóre vo vzťahu k funkčnej úrovni výkonu, na rozdiel od interpretácie odvodenéj od výkonu ostatných osôb“ (s.176). Napr.: porovnanie testových skóre (ich interpretácia) s kritickými skóre (kritické skóre je hodnota na stupnici, na základe ktorej sú skóre nachádzajúce sa pod ňou rozdielne interpretované ako skóre nad ňou).

Štandardy neodporúčajú prístup k chápaniu a zdôvodňovaniu validity na základe typov validity. Adekvátna cesta nevedie cez obsahovú, prediktívnu, konštruktívnu, kritériovú, či inú validitu. V Štandardoch sa zdôrazňuje, že validita je jedna s viacerými jej aspektami. Na rozdiel od minulosti ostatné Štandardy nespájajú priamo validitu s testom, ale precizujú ju ako mieru adekvátnosti **interpretácie** skóre získaných pri administrácii testu. Slovnému spojeniu „validita testu“ sa dôsledne vyhýbajú.

Argumentácia v prospech validity (podľa súčasných Štandardov..) je najčastejšie založená:

- na obsahu testu,
- na vnútornej štruktúre testu,
- na vzťahu k iným premenným,
- na postupoch reagovania na testové položky,
- na dôsledkoch testovania.

(Nazdávame sa, že pri preklade takto zameraných anglických testov do slovenčiny alebo do češtiny dochádza často k závažnej nepresnosti, keď sa anglický výraz **evidence** prekladá ako **dôkaz**, hoci sa tým v danom kontexte myslí termín **podpora, argument v prospech**, alebo jednoducho **evidencia**, čím dochádza k výraznému posunu významu.)

### Evidencia v prospech validity, založená na obsahu testu

V širšom zmysle sa pod obsahovou evidenciou o validite rozumie obsah „rôznych zložiek situácie testovania – predovšetkým položiek testu, ale i obsahu situácie alebo kontextu testovania a obsahových zložiek chovania testovaného v tejto situácii“ (Urbánek ...s. 133) Iné vymedzenie obsahovej evidencie validity ponúkajú Štandady... (s. 19). Poznatky tohto druhu sú získané z analýzy vzťahu medzi obsahom testu a konštruktom, ktorý definuje meranú vlastnosť. Konštrukt je teoretický koncept vlastnosti, ktorú má test merať. Koncept je teda teoretickým obrazom určitej oblasti. Skúma sa 1. miera, v ktorej obsah testu pokrýva oblasť konštrukt a 2. miera relevancie danej oblasti vzhľadom k navrhovanej interpretácii testových skóre (vhodnosť obsahovej oblasti je závislá na konkrétnych zámerov, ktoré majú byť testovaním naplnené). Z takejto pozície sa vyberajú a posudzujú položky testu. Experti posudzujú test ako celok i položku po položke a vyjadrujú sa k miere reprezentatívnosti a relevantnosti použitých položiek. Posudky položky od viacerých expertov je možné kvantifikovať a zmerať tak vhodnosť položky (index skupinového súhlasu s položkou). Obdobne možno postupovať ak jeden expert posudzuje viacero položiek (index príbuznosti položiek z pohľadu jedného experta).

### Evidencia v prospech validity založená na vnútornej štruktúre testu

Štruktúrna (empirická) analýza testu ukáže ako jeho vnútorná štruktúra korešponduje s teoretickou štruktúrou konštrukt. Konštrukt môže byť jedno- alebo viacdimenzionálny. Pre validitu je dôležité, do akej miery vzťahy medzi položkami testu zodpovedajú pojmovému rámcu konštrukt. Ak je konštrukt jednodimenzionálny, položky testu by mali

byť homogénne. Nezriedka, však, testujeme vlastnosti, ktoré sú zložené z viacerých komponentov. Túto skladbu nazývame faktorovou štruktúrou (faktory a dimenzie sa tu používajú ako synonymá). Interpretácia testového skóre, respektíve interpretácia skóre dimenzií testu sa opiera o evidenciu založenú na multidimenzionalite.

Argumentácia v prospech validity tu vychádza z odpovede na dve otázky:

- Je interpretácia skóre testu pomocou vhodného konštruktu validná?
- Produkcia skóre pomocou daného inštrumentu je najlepšia?

Získaná evidencia tohto typu je úzko spätá s existujúcou teóriou, konštrukciou hypotézy odvodenéj z teórie a empirických výsledkov, ktoré korešpondujú s touto teóriou. Nestačia, však, výsledky jediného empirického zisťovania, a to z rovnakého dôvodu, ako nie je možné jedným výskumom dostatočne potvrdiť teóriu. Čím viac máme výsledkov, ktoré potvrdzujú validitu interpretácie testu, presvedčenie o platnosti takejto interpretácie rastie. Psychológia je plná konštruktov, ktoré zobrazujú skryté vlastnosti, a ako také nemôžeme pozorovať bezprostredne.

Tieto konštrukty sú hypotetické a na ich potvrdenie si vyžaduje uskutočniť špeciálne meranie. Výskum, v ktorom sa uskutočňuje určitým spôsobom meranie určitého konštruktu, a ktorý dáva vopred formulované (predikované) výsledky, potvrdzuje nielen samotný konštrukt, ale aj inštrument, ktorým ho meriame.

### **Evidencia v prospech validity založená na vzťahu k iným premenným**

Môžeme na základe výsledkov testovania predvídať budúce správanie? Súvisí toto testovanie s inými meraniami správania? Hovoríme o evidencii založenej na vzťahu testu k iným premenným, pretože výsledky testovania dávame do vzťahu s niektorým kritériom.

Napríklad, kritériálnymi premennými môže byť:

- budúce hodnotenie v škole,
- odhady pomocou iných testov inteligencie.

Aby interpretácia testu úrovne inteligencie bola validná, mal by, napríklad:

- výsledok testu umožniť predvídať úspešnosť žiakov v škole,
- výsledky testu inteligencie by mali byť podobné výsledkom iných testov, ktoré merajú inteligenciu.

Evidencia validity testu založená na vzťahu k iným premenným (včítane vzťahu medzi testom a kritériom – kritériálna evidencia) pokrýva okrem iného obsah toho, čo sa tradične nazýva „kritériálnou validitou“.

Vzťah medzi testom a kritériom možno riešiť takzvanou prediktívnou a súbežnou validizáciou. „Prediktívne štúdie zisťujú, ako presne môžu výsledky testu predpovedať skóre v kritériálnej premennej získané v neskoršej dobe. V súbežnej štúdií meriame prediktor a kritérium v rovnakej dobe“ (s. 22) V minulosti sa v takejto situácii používali termíny „prediktívna, resp. prognostická validita“ a „súbežná, resp. diagnostická validita“.

V skúmaní vzťahov medzi testovými skóre a inými metódami merania sa hľadajú tzv. konvergentné a diskriminačné evidencie. Vzťahy medzi testovými skóre a metódami, ktoré merajú odlišné konštrukty poskytujú diskriminačnú evidenciu. Vzťahy medzi testovým skóre a výsledkami merania podobných konštruktov poskytujú konvergentnú evidenciu.

Evidencia v prospech validity založená **na postupoch reagovania na testové položky** vychádza z teoretickej a empirickej analýzy „spôsobu, ktorým testované osoby odpovedajú na testové položky“. Obvykle sa analyzujú jednotlivé odpovedi (spôsoby odpovedí) testovaných osôb, ich stratégie riešenia testových úloh.

Evidencia v prospech, či neprospech validity založená **na dôsledkoch testovania** sa opiera predovšetkým o výsledky analýzy zamýšľaných a nezamýšľaných dôsledkov použitia testu.

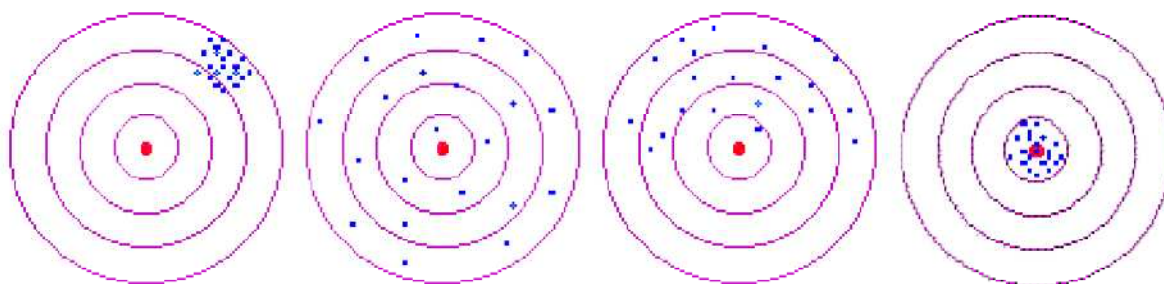
### III. Reliabilita a validita

Aby testovanie malo pre diagnostiku alebo výskum význam, musí byť reliabilné a mať validnú interpretáciu testových hodnôt.

Reliabilita pôsobí na presvedčenie o tom, že uskutočnené meranie poskytuje skóre, ktoré sú blízke pravdivým hodnotám. Validita ukazuje, že interpretáciu testových skóre môžeme považovať za platnú.

Validita predpokladá reliabilitu. Naopak to neplatí. Reliabilné dáta môžeme interpretovať neadekvátne, nevalidne (pozri obr. A), alebo adekvátne, validne (obr. D). Čiže len interpretácia reliabilných výsledkov merania *môže* byť validna, ale je tu možnosť, že nemusí byť validna. Interpretácia nereliabilných (skóre nie sú blízke pravdivým hodnotám) dát (obr. B alebo C) nie je v žiadnom prípade validna.

Otázka reliability a validity má tiež etickú stránku. Zvlášť ak výsledky testovania sa použijú v riešeníach, ktoré sa dotýkajú života ľudí. Uchádzačov o prácu do zamestnania zoberú alebo nie, študentov hlásiacich sa na ďalšie štúdium prijme do školy alebo nie. Od určenia diagnózy pacienta závisí spôsob liečby. Len dostatočná validita a reliabilita testu pomôže urobiť v týchto vážnych situáciách adekvátne rozhodnutia.



Obr. A

Obr. B

Obr. C

Obr. D

Je potrebné si uvedomiť, že pojem validity v tomto texte sa rozoberá v kontexte merania. O validite hovoríme nielen v rámci témy merania nejakej psychosociálnej vlastnosti. Pojem validita sa používa aj ako charakteristika celého sociálneho výskumu. Validita charakterizuje platnosť celého experimentu, alebo proces empirického overovania nejakej teoretickej hypotézy. To je však už iná téma.

## Literatúra:

1. AERA, APA, NCME (1999). Standards for Educational and Psychological Testing. Washington, DC.
2. AERA, APA, NCME (2001). Standardy pro pedagogické a psychologické testování, Praha, Testcentrum.
3. APA (1985). Standards for Educational and Psychological Testing. Washington, DC.
4. BLAHUŠ, P. (1996). K systémovému pojetí statistických metod v metodologii empirického výzkumu chování. Praha. Univerzita Karlova, Karolinum.
5. BRZEZIŃSKI, J. (2003). Metodologia badań psychologicznych. Warszawa, Wydawnictwo naukowe PWN.
6. ČERNÝ, V., KOLLÁRIK, T. (1986). Kompendium psychodiagnostických metod. Bratislava, Psychologické a didaktické testy, n. p.
7. FERJENČÍK, J. (2000). Úvod do metodologie psychologického výzkumu. Praha, Portál.
8. GOODWIN, C., J. (2002). Research in psychology. Methods and design. New York, John Wiley and Sons, Inc.
9. HALAMA, P. (2005). Princípy psychologickéj diagnostiky. Trnava, FF TU.
10. HENDL, J. (2004). Přehled statistických metod zpracování dat. Praha, Portál.
11. HIRNER, A. (1976). Ako sociologicky analyzovať. Bratislava. Ústav školských informácií.

12. HRABOVSKÝ, M. (1990). Špecifickosť štúdia psychických javov a princípy získavania kvantitatívnych údajov o psychologických premenných. In: Maršálová, L., Mikšík O. a kol. Metodológia a metódy psychologického výskumu, Bratislava SPN, ss:109-160.
13. KERLINGER, N., F. (1972). Základy výskumu chováni. Praha, Academia, ČSAV.
14. KOŽENÝ, J. (1998). Teorie, data a validita měření konstruktů. Československá psychologie, roč. 42, č.6, ss:481-502.
15. MARŠÁLOVÁ, L. (1978). Metodologické základy psychologického výskumu. Bratislava, Psychologické a didaktické testy, n. p.
16. MELKA, A. (2005). Základy experimentální psychoakustiky. Praha, Akademie múzických umění.
17. MICHELL, J. (1999). Measurement in Psychology. Cambridge: Cambridge University Press.
18. NOWAK, S. (1975). Metodologie sociologických výzkumů. Praha, Nakladatelství Svoboda.
19. REITEROVÁ, E. (2003). Základy psychometrie. Olomouc. FF Univerzita Palackého.
20. RITOMSKÝ, A. (2002). Metódy psychologického výskumu: kvantitatívna analýza dát. Bratislava, Medzinárodné stredisko pre štúdium rodiny.
21. ŘÍČAN, P. (1977). Úvod do psychometrie. Bratislava, Psychologické a didaktické testy, n. p.
22. SRNEC, J. (2010). Základy psychognostiky. Praha, PVŠPS.
23. STEVENS, S., S. (1946). On the theory of scales of measurement. Science 103 (2684): 677-80.
24. SCHENK J., HRABOVSKÁ A. (2010). Škálovanie. Bratislava, Univerzita Komenského.
25. THURSTONE, L.,L. (1959). The Measurement of Values. Chicago: The University of Chicago Press.
26. URBÁNEK, T., DENGLEROVÁ, D., ŠIRUČEK, J. (2011). Psychometrika. Měření v psychologii. Praha, Portál.

---

## Validity, Reliability and Errors of Measurement

### Abstract

The use of reliability and validity are common in quantitative research. For each total score, subscore, or combination of scores that is to be interpreted, estimates of relevant reliabilities and standard errors of measurement or test information functions should be reported. We perform reliability analyses on the overall score, if applicable, and all subscores. The analyses performed are Cronbach's Coefficient Alpha, and the following split half tests: Correlation between forms, Spearman-Brown, and Guttman's formula. The Standard Error of Measurement scores will soon be added to all research (statistical) reports. As described in the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999) validity refers to "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests". The validation process, therefore, involves the accumulation of evidence to support the proposed test score interpretations and uses (evidence of validity based on test content/content validity, on test-criteria relationships and validity based on a test's relation to constructs). The process of accumulating evidence to support the validity of test score interpretations starts prior to the development of an assessment.

---

**Key words:** Construct, Validity, Reliability, Errors of Measurement, test score interpretations