

Validita diagnostického nástroje

PSYb2590: Základy psychometriky | Přednáška 3

14. 3. 2022 | Hynek Cígler

IDS: Co to asi měří?

Řekněte: **„Teď ti vždycky přečtu čtyři slova a ty z nich máš vybrat to, které se k těm třem ostatním nehodí, které k nim nepatří. Nejdřív jenom dobře poslouchej, já ta slova potom ještě jednou zopakuji a teprve pak vybereš to, které se k ostatním nehodí.“** Zadání každé otázky je možné několikrát opakovat, dítě má za úkol jakýmkoliv způsobem (slovy, gestem ap.) vybrat správnou odpověď.

Pomůcka pro hodnocení: Dítě může odpovídat slovy, ale také jen neverbálně – dát signál jakýmkoli pro něj dostupným způsobem. (Správné odpovědi jsou označeny **tučně**).

„Zkusíme si to, poslouchej dobře“: 1. tričko – **míč** – svetr – kabát.

Zadání jedenkrát opakujte a počkejte na odpověď dítěte. Pokud dítě neodpoví nebo odpoví chybně, řekněte: **„Nepatří tam míč. Tričko, svetr i kabát nosíme na sobě, je to oblečení, ale míč ne, proto k nim nepatří. Zkusíme teď další úkol.“**

Pokud dítě podá odpověď, přijměte ji (ať již je správná nebo ne) a řekněte: **„Ano, děláš to dobře. Zkusíme další otázku.“**

Číslo Úkol

2. jablko – švestka – **rohlík** – jahoda
3. svetr – **kapesník** – kalhoty – rukavice
4. myš – slepice – **strom** – had
5. **motorka** – kolo – koloběžka – skateboard
6. **město** – les – louka – hory
7. paprika – **oves** – okurka – mrkev
8. stůl – **lampa** – křeslo – skříň
9. štika – pstruh – **kosatka** – okoun
10. učitel – holič – **soused** – zedník
11. tenis – **běh** – kopaná – házená
12. smrk – růže – **strom** – fialka
13. **auto** – vlak – tramvaj – autobus
14. **mlha** – déšť – kroupy – sníh
15. ležet – **jít** – klečat – stát
16. hladký – drsný – **chladný** – vrásčitý
17. rejsek – **datel** – netopýr – krtek
18. **sklo** – železo – měď – stříbro
19. **morče** – kočka – vlk – kuna
20. **vidlička** – mísa – sklenice – pekáč



Validita

Validita výzkumu = platnost výsledků studie vzhledem ke skutečnosti.

Validita diagnostické metody = platnost měření vzhledem ke skutečnosti.

- APA (aj. moderní zdroje): měření je operacionalizováno jako „skóry“...
- ... a tedy „**validita testových skóru** vzhledem ke skutečnosti“.

Validita metody je tedy *shoda mezi nástrojem a měřeným rysem*.

- Co ale, když měřený rys nedává smysl?
- Jak míru této shody vyjádřit?

E. G. Boring: „Intelligence je to, co měří test intelligence.“

- Upozorňuje na problém *reifikace*.

Historie konceptu validity testu 1

Okolo roku 1915 pojem validity ve smyslu prediktivní validity.

- „Here I am distinguishing between two different but related ideas namely, reliability and validity. An instrument of measurement is reliable to the extent that it yields the same results at different times and in the hands of different persons. **It is valid to the extent that it measures the thing it is supposed to measure.**“ (Buckingham, [1921](#)).

Kodifikace ve třicátých letech (Thurstone, 1931; Guilford, 1936), důraz na souběžnou validitu.

- A na obsahovou (výběrovou) validitu.

Validita byla používána ve smyslu „korelace“ (s kritériem), nový test měl korelovat s předchozím testem.

- Test IQ1 dostatečně koreluje s kritériem K. Nový test IQ2 dostatečně koreluje s původním testem IQ1. Novější test IQ3 dostatečně koreluje s IQ2. Nejnovější IQ4 se starším IQ3. Co měří test IQ4?

Tato pojetí byla neudržitelná, protože neříkala nic o tom, **co** metoda měří.

Podle Lissitz a Samuelsen ([2007](#)).

Historie konceptu validity testu 2

Cronbach a Meehl ([1955](#)): **Construct Validity** in Psychological Tests.

- „Construct validity is ordinarily studied when the tester has no definite criterion measure of the quality with which he is concerned, and must use indirect measures. Here the trait or quality underlying the test is of central importance, rather than either the test behavior or the scores on the criteria.“
- Navázali např. Campbell a Fiske ([1959](#)) s metodou Multitrait-multimethod matrix (MTMM).

Vychází z **logického pozitivismu**.

- Existuje **nomologická síť**¹ teoretických konstruktů, které jsou vzájemně propojeny.
- Tyto konstrukty jsou operacionalizovány do pozorování (testem).
- Odpovídají pozorované vztahy těchto operacionalizací (testů) teoretickým vztahům (konstruktů) tak, jak odpovídá uvažované nomologické síti?

¹ Nebo lépe viz Cronbach a Meehl (1955)

Historie konceptu validity testu 3

Messick (zejm. 1989, [1995](#)) vytvořil zastřešující pojetí konstruktové validity (unified construct validity).

- Toto pojetí je dnes dominantní, převzaly jej i Standardy pro pedagogické a psychologické testování (AERA, APA, NCME, 1999/2014).

V posledních 30 letech probíhá další vývoj v souvislosti s *model-based* psychometrikou.

- **Borsboom, Michell** a jiní. [Borsboom se kriticky vymezuje](#) vůči pojetí konstruktové validity.
- Nová, zastřešující pojetí (např. [Lissitz a Samuelsen, 2007](#)).

Jeden kurz, tři pojetí validity

Psychometrikové se nedokážou shodnout na jediném pojetí validity. Je potřeba znát všechny.

A. Základní „druhy“ validity...

- *Obsahová a empirická*; další klíčové pojmy.

... + tradiční pojetí *konstruktové validity*.

- K předchozím přidává ještě konstruktovou validitu, dohromady jde o tradiční pohled na validitu.
- Cronbach a Meehl (1955).

B. Moderní pojetí *unifikované konstruktové validity*.

- Messick (1989, 1995).

C. Moderní pojetí vycházející z realismu („model-based“).



Tradiční pojetí: Tři zdroje důkazů

Obsahová validita: Shoda mezi metodou a účelem metody (v teoretické rovině).

- Zjevná, zdánlivá, výběrová, obsahová, faktorová...

Empirická validita: Shoda mezi výsledkem testu a vnějším kritériem.

- Konvergentní, divergentní (diskriminační, diferenciální).
- Souběžná, prediktivní (dříve samostatná oblast důkazů), postdiktivní.
- Kriteriační.
 - Někdy synonymum pro empirickou v., jindy termín pro empirickou validitu s kategorickým/objektivním kritériem.
- Inkrementální.

Konstruktová validita: shoda metody s latentními konstrukty postulovanými teorií.

- Zaměřená na latentní proměnné (strukturní modelování, MTMM), faktorová analýza.
- Flexibilní kladení a zodpovídání otázek vyplývajících z účelu metody a teorie.

Moderní pojetí: Unifikovaná konstruktová validita

Vychází z tzv. Messickova (např. [1995](#)) „sjednoceného pojetí konstruktové validity“.

- Validita je jediným, multifasetovým konstruktem.
- Důraz na hodnocení a použití testu.

Validita je integrativním shrnutím dílčích důkazů.

Toto pojetí se vyskytuje v moderních učebnicích psychometricky.

Zdroje důkazů podle Messicka:

- Obsah testu
- Vnitřní struktura testu
- Odpověďové procesy
- Souvislost s kritériem
- Konsekvence testování
- (Zobecnitelnost a „šíře významů“.)

Definice konceptu validity

Messick (1989, s. 20): „... *an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment.*“

- Validita je hodnocení testu (proto i např. konsekvenční validita, viz později).

Borsboom (2004): „A test is valid for measuring an attribute if (a) the **attribute exists** and (b) variations in the attribute **causally produce variation in the measurement outcomes.**“

- Validita je vlastnost testu, kterou hodnotíme.

Borsboomova kritika Messickova pojetí 1

„How construct validity is sold:

Construct validity is an evaluative, integrated judgement of the degree to which test score interpretations are justified in the light of empirical evidence and theoretical rationales (and, possibly, social consequences that follow from test use).“

Borsboomova kritika Messickova pojetí 2

„What construct validity really is:

Somebody's evaluative, integrated and fluctuating judgement of the degree to which test score interpretations, that may have nothing to do with measurement, are justified in the light of time-dependent empirical evidence and that person's theoretical rationales (and, possibly, that person's guesses about social consequences that follow from test use as well as his or her valuation of these outcomes).“

Messick vs. Borsboom

MESSICK

„Psychodiagnostické pojetí“.

Validita je **zhodnocením shody nástroje a účelu diagnostiky.**

Validita zahrnuje externí aspekty diagnostiky včetně **důsledků měření.**

Zaměření na **použití metody** v praktickém kontextu.

Konstruktivismus.

BORSBOOM

„Ontologické pojetí“.

Validita je **shoda nástroje a měřeného atributu.**

Externí aspekty měření a důsledky diagnostiky nejsou součástí konceptu.

Zaměření na to, **co je měřeno.** Kontext je důležitý, ale nejde o validitu.

Realismus.

Validita je tedy odpověď na otázku:

(Borsboom a podlézavě i Cígler)

„Je naše teorie o odpovídání lidí správná?“

Dvě dílčí odpovědi:

- **Existuje atribut?**
 - V operacionalistické CTT pravý skór (ne)existuje vždy; CTT je celkově vadný model měření. Preference FA, IRT při validizaci.
- **Měří metoda tento atribut?**

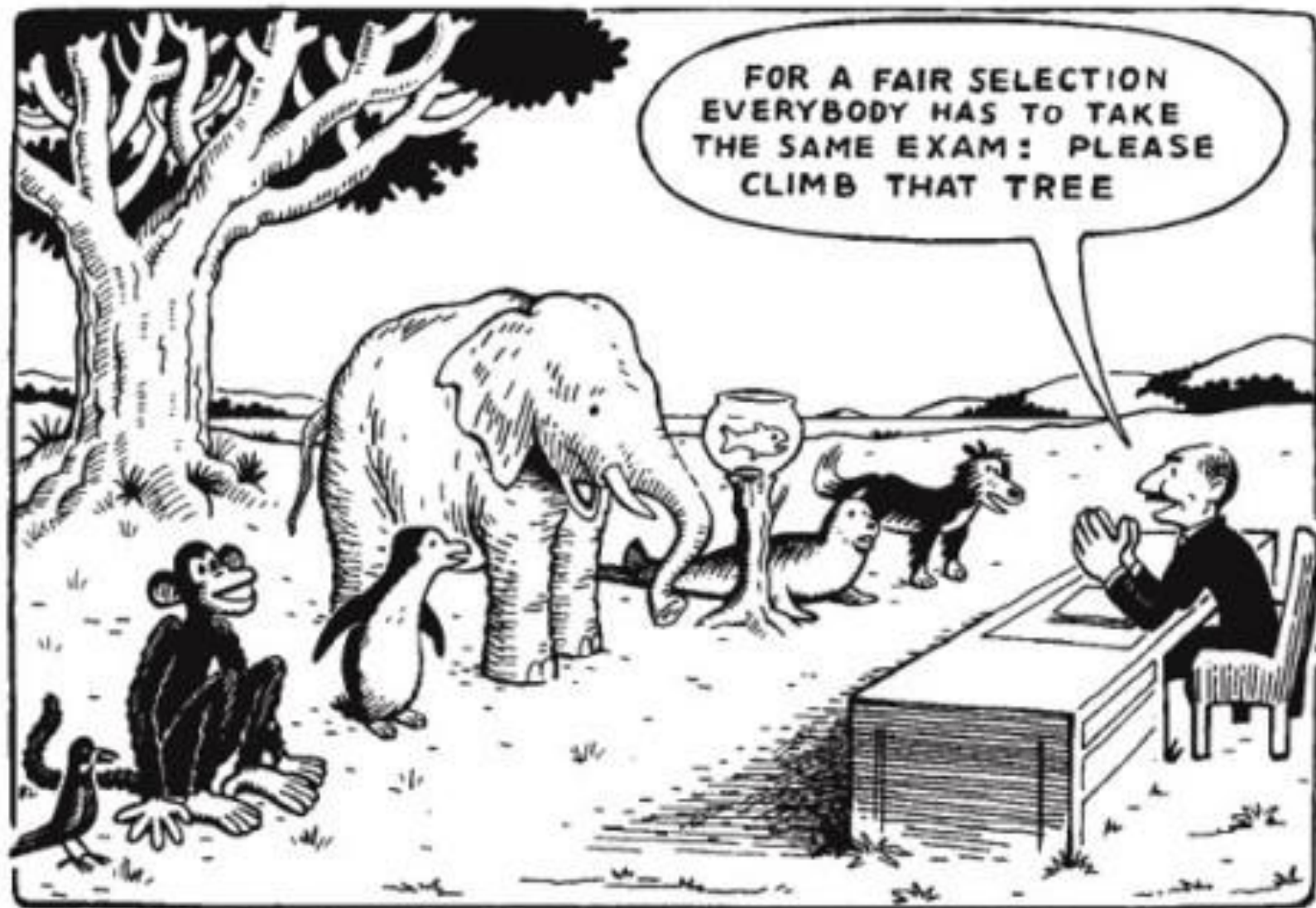
Pod souhrnný pojem „validita“ jsou nicméně často řazeny odpovědi na další důležité otázky:

- Predikuje test chování osob v budoucnu?
- Je použití testu eticky a legálně obhajitelné?
- Slouží použití testu zamýšlenému účelu?

Vybrané pojmy a problémy spojené s konceptem validity

Následují výhradně vybraná témata.

Pro celkový přehled je nezbytné
samostudium.



Zdroje „invalidity“

PODREPREZENTOVÁNÍ KONSTRUKTU

Metoda neměří požadovaný konstrukt v celé šíři.

Důležité aspekty konstruktů nejsou metodou zachyceny.

Typickým důsledkem je psychometrický paradox.

KONSTRUKTOVĚ-IRELEVANTNÍ ROZPTYL

Metoda měří i něco jiného/dalšího.

Část rozptylu skóreů je způsobena něčím jiným.

Např.: Odpovědní styl respondentů, kulturní/jazykové rozdíly (DIF, non-invariance)...

Nomologická síť

Klíčový aspekt konstruktové validity podle Cronbacha a Meehla (1955).

- Zařazení *významu* metody do sítě veškerých možných jiných významů.

Součásti a aspekty nomologické sítě:

- 1. Na základě teorie předpokládáme existenci **několika konstruktů**.
- 2. Teorie předpokládá **vztahy** těchto konstruktů (kauzální?).
- 3. **Operacionalizace** konstruktů do měřicích nástrojů (korespondence).

Operacionalizované proměnné by měly odrážet vztahy konstruktů.

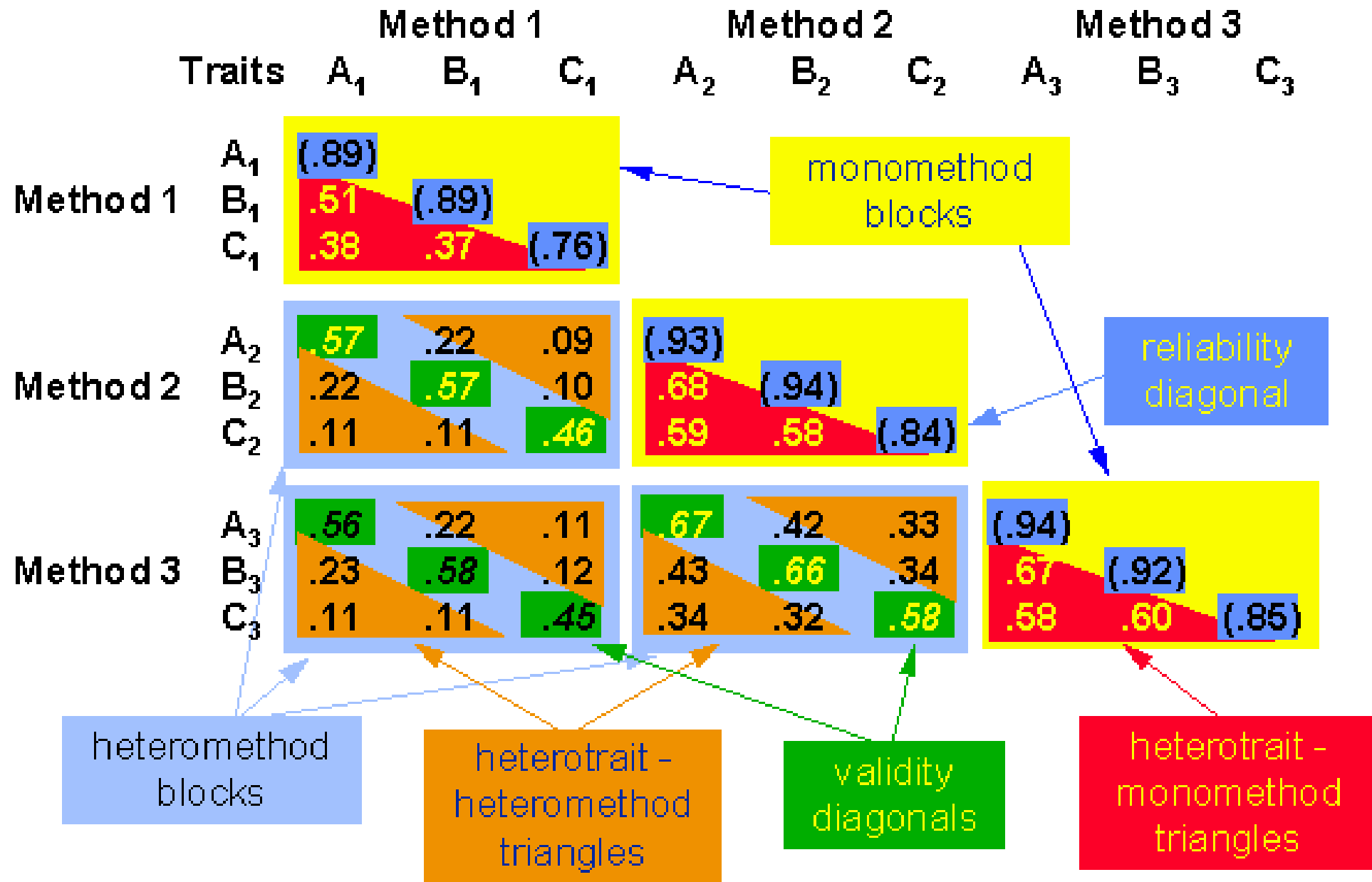
- Pokud ne, metoda není validní, nebo je teorie neplatná.

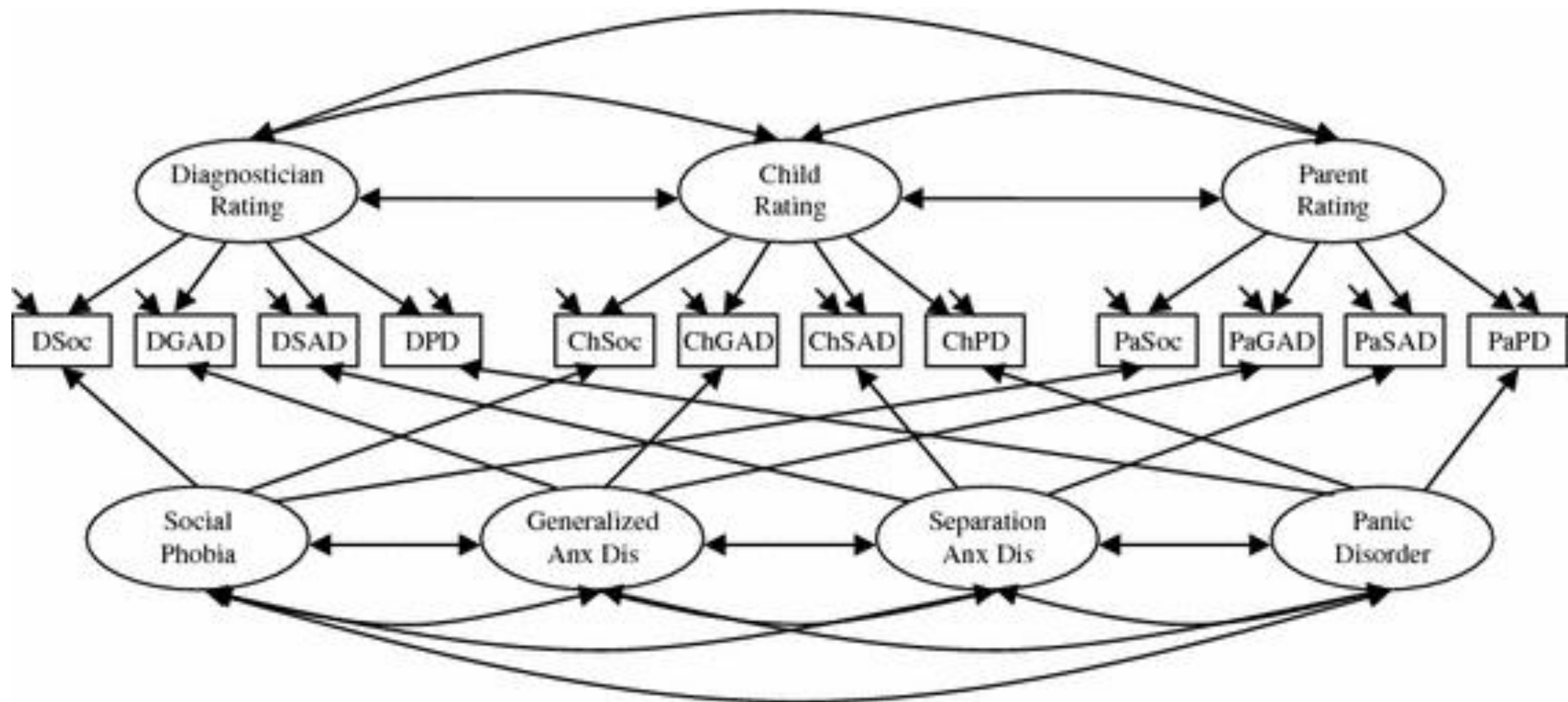
Postupy: Multitrait-Multimethod Matrix (Campbell a Fiske, [1959](#)), SEM model.

TABLE 1
A SYNTHETIC MULTITRAIT-MULTIMETHOD MATRIX

		Method 1			Method 2			Method 3			
Traits		A ₁	B ₁	C ₁	A ₂	B ₂	C ₂	A ₃	B ₃	C ₃	
Method 1	A ₁	(.89)									
	B ₁	.51		(.89)							
	C ₁	.38	.37		(.76)						
Method 2	A ₂	.57			.22		.09				(.93)
	B ₂	.22	.57		.10			.68		(.94)	
	C ₂	.11	.11	.46		.59	.58		(.84)		
Method 3	A ₃	.56		.22	.11		.67	.42	.33		(.94)
	B ₃	.23	.58	.12	.43	.66	.34		.67	(.92)	
	C ₃	.11	.11	.45	.34	.32	.58	.58	.60	(.85)	

Note.—The validity diagonals are the three sets of italicized values. The reliability diagonals are the three sets of values in parentheses. Each heterotrait-monomethod triangle is enclosed by a solid line. Each heterotrait-heteromethod triangle is enclosed by a broken line.



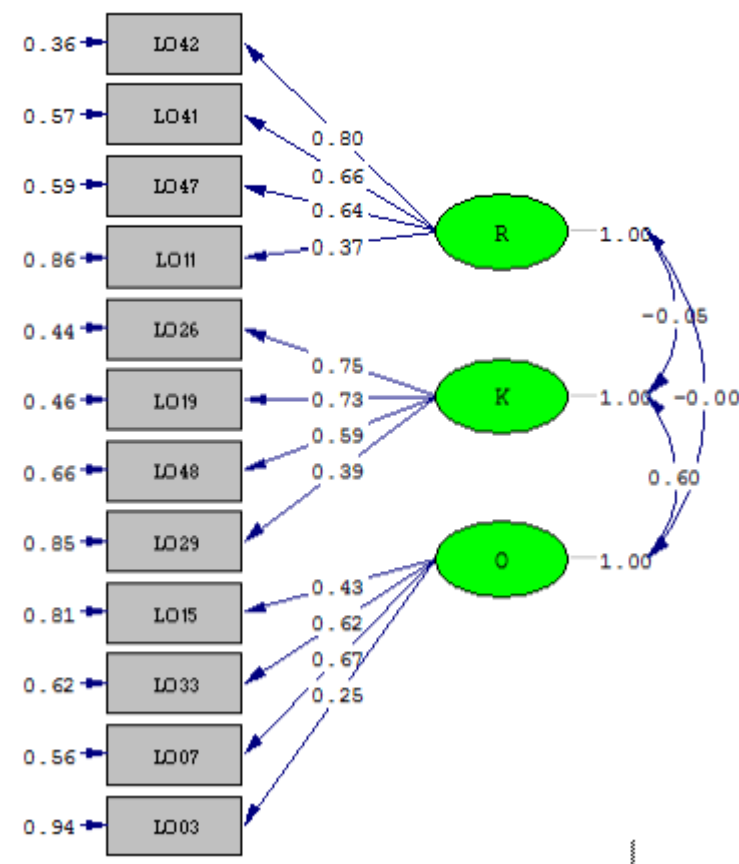


Faktorová validita

Příklad: **Normativní styl identity** (část Berzonskeho inventáře ISI):

- R = ability to achieve closure
- K = potřeba jednat ve shodě s ostatními
- O = potřeba jednat ve shodě se svým přesvědčením

Kód	Text položky	R	K	O
LO11	Vždy jsem věděl/a v co věřím a v co nevěřím; nikdy jsem v tomto směru neměl/a nějaké velké pochybnosti.	X		
LO41	Pokud mám učinit nějaké důležité životní rozhodnutí, většinou nepotřebuji moc času. Brzo vím, jak se rozhodnout.	X		
LO42	Pokud se dostanu do nějakých osobních problémů, celkem hned vím, co mám udělat.	X		
LO47	Když se v mém životě objeví nějaké problémy, řeším je obvykle rychle a bez větších pochybností.	X		
LO19	Myslím si, že je lepší držet se pevně osvědčených názorů, než být příliš svobodomyšlný/á a otevřený/á názorovým změnám.		X	
LO26	Myslím si, že je lepší držet se osvědčených zásad a hodnot, než zkoušet něco nového.		X	
LO29	Mám raději takové situace, v nichž bezpečně vím, jak se mám chovat či jaká v nich platí pravidla.		X	
LO48	Držím se zásad a pravidel, která se už mnohým lidem osvědčila.		X	
LO03	Zásady, ve kterých jsem byl/a vychován/a, jsem přijal/a za své a řídím se jimi.			X
LO07	Snažím se o dosažení těch cílů, které pro mne moji blízcí považují za dobré.			X
LO15	Protože se snažím naplnit očekávání lidí, kteří pro mě hodně znamenají, neřeším problém, jak chci žít a co chci v životě dělat.			X
LO33	Když se rozhoduji o své budoucnosti, často se nechávám bezděčně vést tím, co ode mne očekávají moji blízcí.			X



Inkrementální validita

Máme dva testy (T1, T2) a kritérium (K), které chceme predikovat.

„Zlepšuje souběžné použití T1 a T2 predikci K oproti použití pouze T1?“

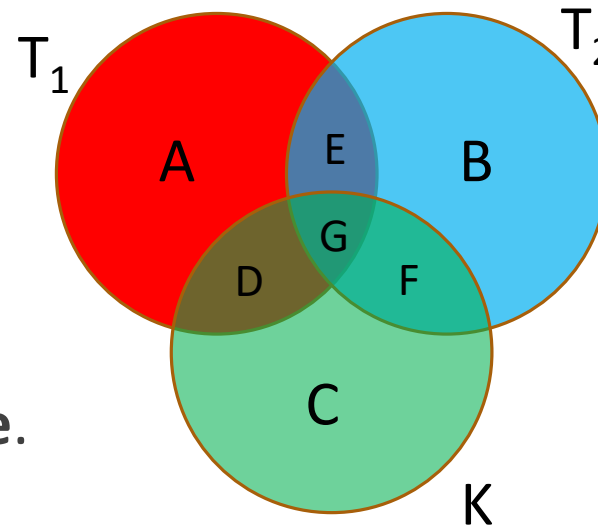
- Např.: „Přijímací zkoušky na FSS: Přispívá test ze ZSV k predikci úspěšnosti u bc. státnic oproti predikci na základě pouze OSP?“

Metoda 1: Parciální korelace.

- Koreluje spolu T2 a K po kontrole T1?
- Koreluje spolu unikátní rozptyl T2 a K, který není společný s T1?

Metoda 2: Hierarchická lineární regrese.

- Krok 1: $K \leftarrow T1$; R^2 vysvětleného rozptylu.
- Krok 2: $K \leftarrow T1 + T2$; $R^2 + \Delta R^2$ vysvětleného rozptylu.
- ΔR^2 je ukazatelem inkrementální validity. Zvýší přidání metody T2 vysvětlený rozptyl kritéria K?



$$r_{T_1K|T_2} = \frac{r_{T_1K} + r_{T_1T_2}r_{T_2K}}{\sqrt{(1 - r_{T_1T_2}^2)(1 - r_{T_2K}^2)}} = \frac{D}{D + A}$$

$$r_{T_1K} = \frac{(D + G)}{(D + G) + (C + F)} = \frac{(D + G)}{(D + G) + (A + E)}$$

Obtíže s kritériem: Nereliabilita

Kritérium, které používáme pro validizaci, není většinou zcela reliabilní (viz attenuation formula).

Dochází k podhodnocení validity.

- Z toho důvodu preference SEM modelů, které odhadují přímo vztah latentních proměnných.

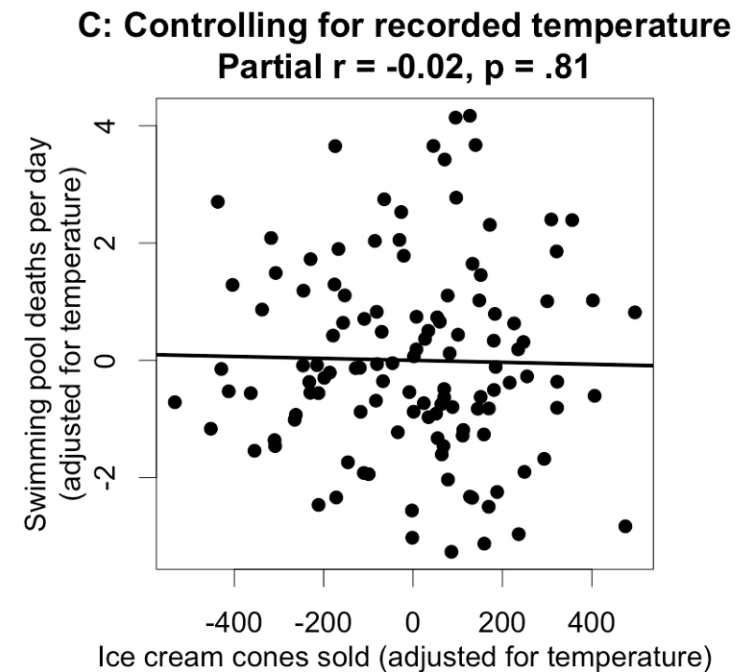
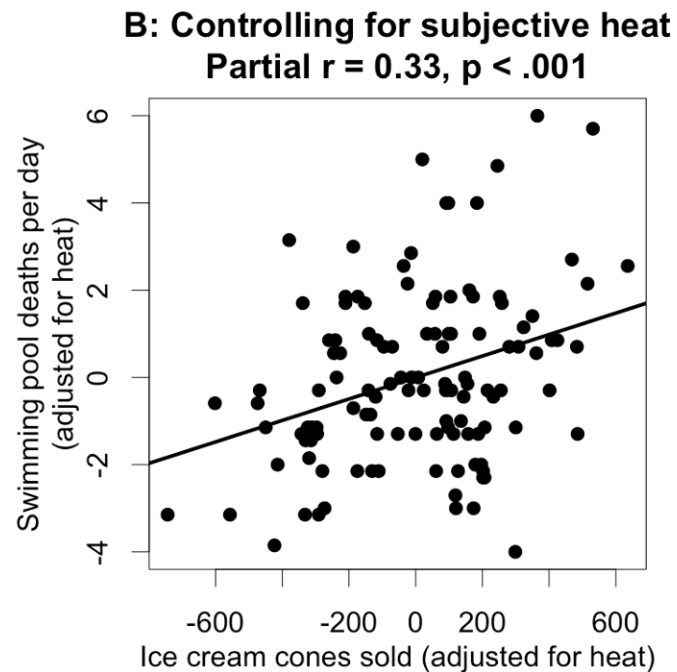
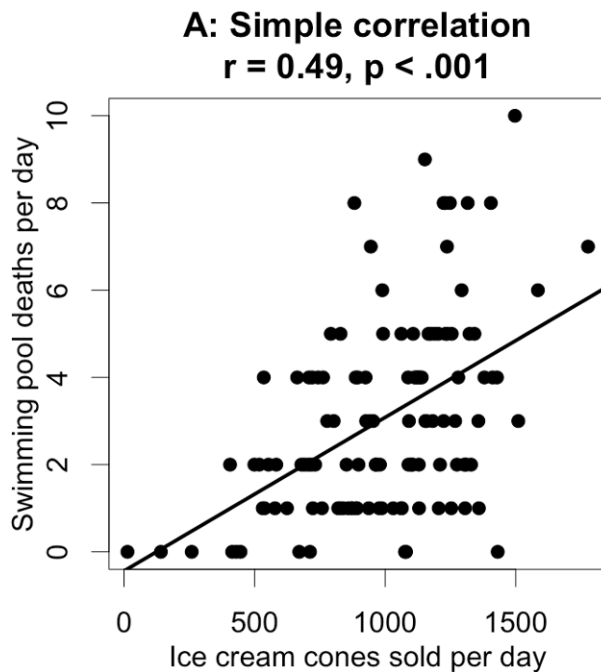
V některých případech ale (ne)reliabilitou *dělíme* – např. parciální korelace, inkrementální validita.

- V takovém případě nereliabilita kritéria může **nadhodnotit** odhad konstruktové validity.
- Nová metoda zvyšuje predikci, ale může vést k výzkumně vadným závěrům.

Obtíže s kritériem: Nereliabilita

Vztah počtu utonutí a prodeje zmrzliny (na základě prodeje zmrzliny predikujeme, kolik lidí se ten den utopí po kontrole teploty).

- Westfall a Yarkoni ([2016](#)).



Obtíže s kritériem: Distribuce

Zejména při ověřování prediktivní validity často pozorujeme efekt stropu či podlahy.

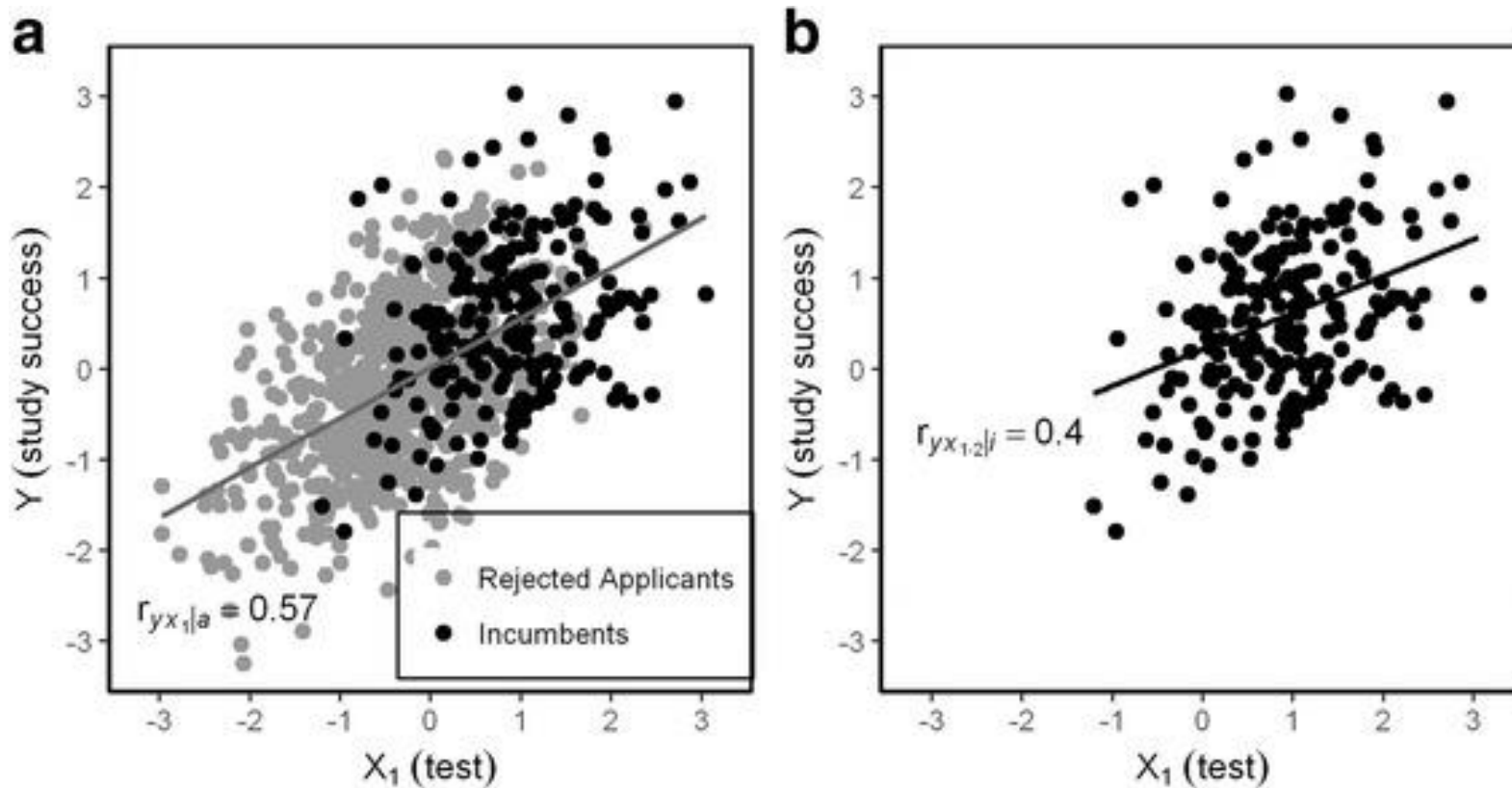
- Na základě testu je přijato 200 studentů z 1000 uchazečů o studium na VŠ.
- U těchto 200 studentů je ověřena prediktivní validita testu vůči prospěchu na konci 1. ročníku.

Efekt stropu/podlahy snižuje rozptyl a tedy pozorované korelace: validita je podhodnocena.

- Existuje řada postupů pro korekci těchto efektů.

Může docházet ke zdánlivým korelacím.

Omezení rozptylu (efekt stropu/podlahy)



The correlation in the full sample of applicants (a) is larger than the correlation in the incumbents

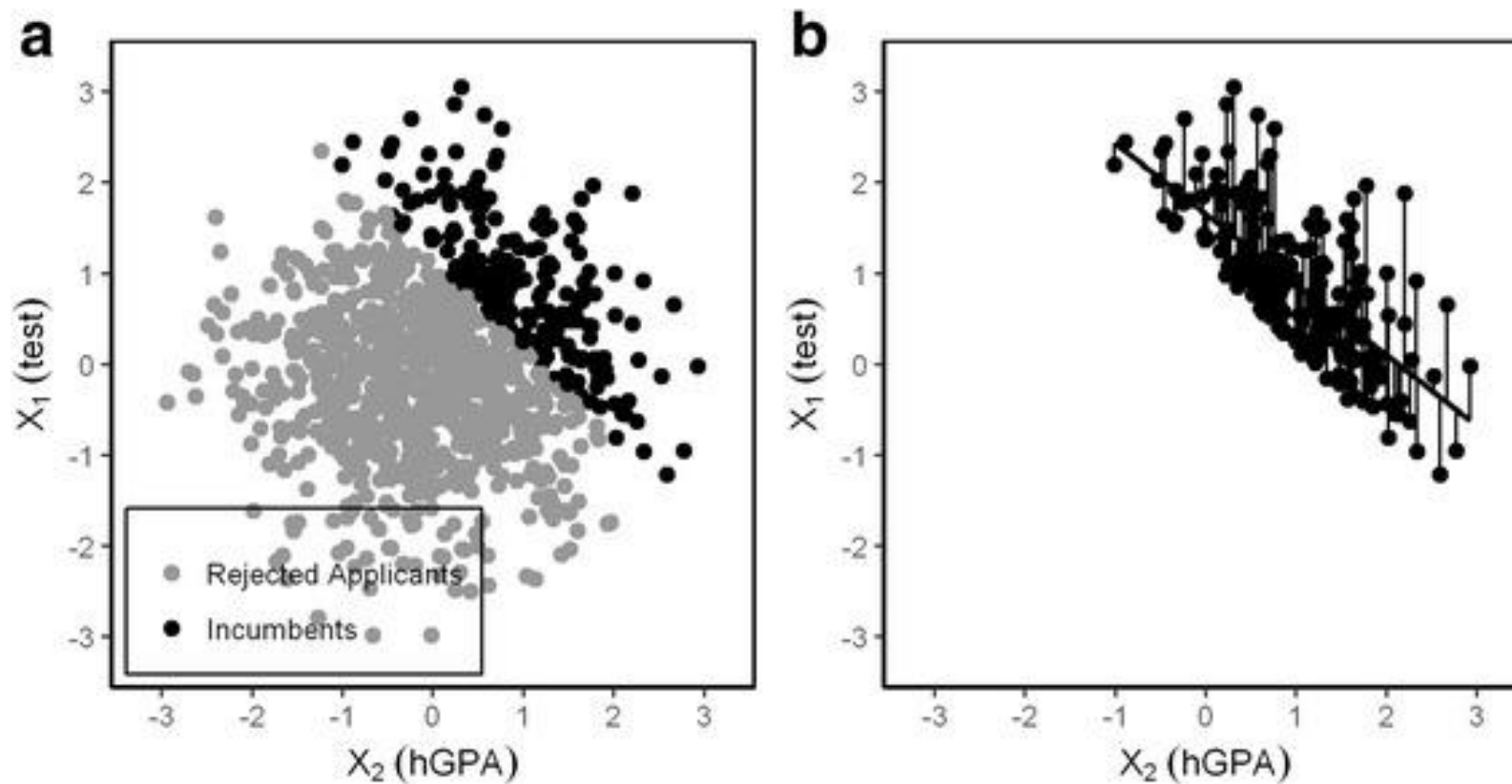
(b) due to range restriction: The variances of X_1 , X_2 and Y are restricted

Zimmerman a kol. (2017)

Thorndikeovy rovnice.

V češtině např. dizertace Viktorové (2018).

Kompenzatorní efekt



a) Scattergram of X_1 and X_2 . 20% of 1000 applicants are selected by the sum of X_1 and X_2 ; the circular cloud representing all applicants is divided by a diagonal line that separates the top right area from the bottom left area.

b) This generates a negative correlation between X_1 and X_2 in the incumbents ($r_{x_1x_2|i} = -0,71$). Residuals of X_1 after the linear effect of X_2 is removed. They are expressed as deviations from the regression line: The residuum of X_1 when the influence of X_2 is removed is the observed X_1 value minus the expected value of the regression X_1 on X_2 .

Zimmerman a kol. (2017)

Messick 1

1. OBSAHOVÁ RELEVANCE

Odpovídá obsah testu účelu?

Nechybí něco, nepřebývá něco, co by podle teorie (ne)mělo být součástí testu?

Byl dobrý způsob tvorby položek?

De facto **výběrová validita**.

2. VNITŘNÍ STRUKTURA TESTU

Odpovídá faktorová struktura očekávání na základě teorie?

- Dříve tzv. **faktorová validita** jako součást validity obsahové.

Je pozorovaná faktorová struktura shodná s tou zamýšlenou?

Někdy zahrnuje i zvažení **vnitřní konzistence**.

- Je vnitřní konzistence/reliabilita taková, jaká by měla být?

Messick 2

3. ODPOVĚĎOVÉ PROCESY

Volí při odpovídání respondent takové kognitivní procesy, které by na základě teorie volit měl?

Neintervenuje nějaký externí systematický rozptyl?

- Odpověďový styl, sociální žádoucnost.

Rorschach, projektivní testy...

Rozumí respondent položkám?

4. SOUVISLOST S DALŠÍMI PROMĚNNÝMI

De facto původní pojetí konstruktové validity podle Cronbacha a Meehla (1955).

Souhrn důkazů konvergentní a divergentní validity, zařazení skóru do nomotetické sítě.

Úzce souvisí s tradičním pojetím empirické validity.

Messick 3

5. KONSEKVENCE TESTOVÁNÍ

Úzce souvisí s férovostí testování.

- Nezvýhodňují přijímací zkoušky muže na úkor žen? Majoritu na úkor Romů?

Culture-free testing.

Nejsou výsledky testování zneužitelné?

6. ZOBECNITELNOST SKÓRŮ

Jaká je zobecnitelnost testových výsledků, jak široký je jejich význam?

Co všechno ještě „*znamená*“ určitý výsledek v testu a co už ne?

Validizační zjištění se netýkají výzkumného vzorku, ale konstruktů jako takového.

Podobné ekologické validitě ve výzkumu.

V novějších úpravách modelu je vyřazeno.

- Součástí vnitřní struktury, resp. zobecnitelnosti skóre.

Messick: Konsekvence testování

Lissitz ([2009](#)): Ilustrativní příklad.

- Máme běžný test školní zralosti, který dobře predikuje neúspěch v prvním ročníku.
- Dítě, které jej nezvládne, není *školně zralé*. Předpokládáme, že jde o zpožděný vývoj.
 - A tedy předpokládáme, že díky odkladu školní docházky dítě své vrstevníky „dožene“.
- Co když ale dítě pochází ze znevýhodněného, nepodnětného prostředí?
 - Vývoj tedy není zpožděný pomalejším zráním, ale nedostatečnými příležitostmi k učení.
- Odklad potíže naopak prohloubí, protože dítě v nepodnětném prostředí zůstane i nadále.
- Test validně odliší školně zralé a nezralé děti. Naše rozhodnutí na základě validních výsledků již validní není, konsekvence testování neodpovídají našim cílům.

Kauza D. H. a ostatní versus ČR ([2007](#)): Reálný tuzemský příklad [[1](#), [2](#)].