06

Original file

6) Konstrukce a vlastnosti psychodiagnostických metod. Psychometrické vlastnosti testů. Standardizace. Položková analýza. Validita diagnostické metody a způsoby jejího zjišťování, druhy validity. Reliabilita a způsoby jejího zjišťování, zdroje chyb měření. Normalizace a druhy norem (věkové, kvantilové, standardní)

Konstrukce psychodiagnostické metody spočívá nejčastěji v tvorbě dostatečného počtu vhodných položek, mezi kterými je poté nutné vybrat ty, které jsou z hlediska relevantní teorie a na základě empirických dat vhodné pro měření daného atributu. Provádění takového výběru probíhá za pomoci nějaké metody analýzy položek.

Pokud tvůrce metody vyjde z toho, jaký atribut by měla jeho metoda měřit a u jaké skupiny porobandů, měl by nejprve vymezit a rozdělit příslušnou obsahovou doménu (obsahové univerzum) podle důležitých vlastností nebo dimenzí měřeného atributu tzv. **faset** (viz. psychometrika str. 205 – 207 – fakt nevím jak to popsat, ani jestli to je důležité, tak se na to radši mrkněte, tu knížku mám když tak naskenovanou). Reprezentativním výběrem témat ze všech těchto faset se vlastně tvůrce snaží naplnit požadavky **obsahové validity.**

1. liší se podle toho, zda se snažíme zjišťovat úroveň nějaké **kognitivní schopnosti** (dovednosti, vědomosti atd.) nebo **osobnostního rysu.**
2. U položek pro měření **kognitivních rysů** jde vlastně o vytvoření sady úkolů, které by měli lépe řešit lidé s vysokou mírou schopnosti než lidé s nízkou mírou schopnosti.
3. Položek by měl být dostatečně velký počet a měly by se dostatečně lišit z hlediska obtížnosti, aby byl test dostatečně citlivý a rozlišoval mezi probandy s různými úrovněmi schopností
4. U dotazníkových metod pro **měření úrovně nějakého rysu osobnosti** je situace o něco složitější, protože na osobnostní rys se nelze zeptat přímo (resp. dá, ale odpověď, kterou bychom dostali, by se jen stěží dala považovat za validní)
5. na míru nějakého rysu můžeme usuzovat pouze z odpovědí na otázky (samozřejmě pokud pomineme některé metody klinické psychodiagnostiky, z nichž nejklasičtější je pozorování)
6. např. Nepřímé otázky: Myslíte si, že je důležité chodit včas? Nebo přímé otázky: Jste dochvilný? Položka může mít také formu započatého výroku k dokončení.

Měli byhom se vyhýbat složitým nebo nezvyklým formulacím (např. Také dvojitý zápor ve větě, který se v češtině často vyskytuje)

Všechny nesprávné varianty by měly být přibližně stejně přijatelné

Počet alternativ větší než dvě.

Neměly by se oběvovat překrývající se nebo navazující odpovědi na položky

U neúplné věty, pro kterou se má vybrat dokončení, je nutné, aby alternativy správně gramaticky vyrůstaly z hlavní věty

U testů kognitivních schopností by odpovědi mely být podobné délky a správné odpovědi by neměly být na určitém místě častěji než jindy

Formáty odpovědí

1. *Doplňování odpovědi*

7+2=…; Co mají společného obraz a socha?; Za dvacet let budu…

- *Vícenásobná volba*

Výběr z více než 3 variant s 1 správnou odpovědí a x distraktory

1. *Dichotomická položka*

Výběr 2 variant

1. *Mnohonásobná volba*

Vlastně série dichotomií…

1. *Spojování (párování)*

„uspořádané kategorie“ resp. „stejně se jevící intervaly“

Nikdy –zřídka –občas -často –vždy

1. *Nucená volba*

Jste spíše tolerantní, nebo náročný?

1. *Řazení*

Seřaďte výroky podle výstižnosti: Jsem energický… nezávislý … důvěřivý…

1. *Neuspořádané kategorie (vícenásobná volba bez kriteriální odpovědi)*

Rozlišovací účinnost položky

Pokud má nástroj měření fungovat, musí především rozlišovat mezi měřenými objekty z hlediska atributu, který má měřit. Asi těžko bychom chtěli měřit tloušťku různých tenkých drátů pomocí obyčejného kancelářského pravítka, protože citlivost takového měření by byla velmi pochybná. Protože položky považujeme za základní prvky, pomocí kterých měření probíhá, musí být každá jednotlivá položka dostatečně citlivá z hlediska měřeného rysu. To znamená, že nežádoucí jsou takové položky, na které všichni odpovídají stejně, nebo úkol, který všichni bez problémů vyřeší (nebo který není schopen vyřešit nikdo). Samozřejmě je nutné uvažovat také o tom, pro koho je metoda určena; např. metoda měřící sílu nějakého postoje bude necitlivá v případech, kdy respondenty budou osoby vyznačující se extrémními hodnotami tohoto postoje.

Analýza položek

Tato analýza probíhá tak, že se pokusná verze testu předloží vhodnému výběrovému souboru osob a získaná data týkající se relevantních charakteristik probandů (např. Věku, pohlaví, profese, diagnózy) spolu s údaji o tom, jak jednotliví probandi odpovídali na každou jednotlivou položku, jsou statisticky zpracovány.

1. dostatečně velký a reprezentativní výběrový soubor (minimálně dvakrát až třikrát vyšší než počet položek, nebo v případě použití faktorové analýzy dvakrát až třikrát vyšší než počet parametrů modelu)
2. výběrový soubor by se měl svým složením co nejvíce podobat složení cílové populace
3. v ideálním případě náhodný výběr

Metody analýzy položek

Přístup založený na kritériu (maximalizaci vztahu s kritériem)

1. Z historického hlediska nejstarší
2. Cíl - vytvořit takovou metodu, která bude spolehlivě rozlišovat mezi určitými skupinami osob nebo která bude nejlépe predikovat nějaké kritérium
3. Skupiny mohou být definovány např. jako diagnózy. Příkladem metod, které byly vytvořeny pomocí tohoto postupu a jsou dodnes s úspěchem používány, je např. metoda MMPI (Minnesotský multi-fázový osobnostní dotazník) nebo Strongův zájmový inventář
4. porovnávají se průměry dvou skupiny, ale analýzu je zejména nutné provést na úrovni položek – hlavně je třeba zjistit, které skóry položek nemají statisticky významný vztah s kritériem – takové s největší pravděpodobností nepřispívají ke vzniku zjištěného rozdílu v naměřeném skóru obou skupin.
5. Hlavní nedostatek této metody spočívá v tom, že přístup k analýze položek je většinou čistě empirický a není korigován žádnou teoretickou koncepcí, která by napomáhala vysvětlení výsledků zjištěných analýzou
6. Někteří autoři (Helmstadter, 1964) mluví přímo o absenci psychologického významu - tvrdí, že zkoumané skupiny se od sebe často neliší pouze z hlediska rysu měřeného danou metodou, ale i z hlediska jiných charakteristik, takže je vlastně obtížné rozhodnout, co škála měří
7. Položky se vyhodnocují jako dobré nebo špatné podle toho, nakolik korelují s kritériem
8. Problém – aby bylo možné prokázat rozdíly mezi skupinami, je nutné tyto skupiny vhodným způsobem vymezit a vybrat. Klasifikace do skupin ale může mít velmi problematickou reliabilitu, čímž může být celý postup analýzy položek značně znehodnocen

Přístup založený na klasické analýze položek

1. klasická analýza položek je postup, který používá poměrně běžné statistické indexy pro vyjádření charakteristik položek, vztahů mezi položkami navzájem a vztahů položek s celkovým skórem testu
2. Princip je velmi podobný jako v předchozím případě postupu založeného na kritériu - opět je dostatečně velkému a reprezentativnímu výběrovému souboru (uvádí se, že minimální rozsah souboru je N = 100) administrována pokusná verze metody, je provedena statistická analýza a jsou vyhodnoceny výsledky¨
3. Rozdíl je v cíli – minulý postup byl založen na maximalizaci validity, tento na maximalizaci reliability – cílem je tedy vytvořit co nejvíc vnitřně konzistentní test

Obtížnost položky – u výkonových testů - zjištění podílu osob z analyzovaného souboru, které danou položku vyřešily správně – pravděpodobnost zodpovězení pro použitelnou položku by měla

být mezi 0,2 a 0,8

Rozlišovací účinnost položky - míra, do jaké se probandi liší ve svých odpovědích na ni

1. hlavní myšlenkou vysvětlující hodnotu korelace skórů položky a skórů testu jako míru rozlišovací účinnosti položky spočívá v tom, že pokud je tato hodnota dostatečně vysoká, znamená to, že skór položky má dostatečně vysoký podíl variability společná

se skórem celého testu.
2. Položka s dostatečně vysokou rozlišovací účinností by měla mít hodnotu korigované korelace (od skóru testu se odečte skór položky, kterou korelujeme) skóru položky se skórem testu alespoň 0,3.

Přístup založený na faktorové analýze

1. U tohoto přístupu platí naprosto stejné principy a používá se stejný postup jako u klasická analýzy položek
2. Specifičnost tohoto postupu oproti klasické analýze položek spočívá v tom, že se namísto korelační analýzy používá faktorová analýza, a to pokud možno konfrimační faktorová analýza, protože umožňuje testovat model vztahů mezi skóry položek, tzn, kriticky zkoumat nějakou teorii

Pak jsou ještě ještě přístupy založené **na teorii odpovědi na položku (IRT) a na teorii vědomostního prostoru (KST)**. Těm jsou ale věnovány samostatné kapitoly před tím, než se píše o analýze položek. V kapitole věnující se analýze položek o nich nic není ….

Chyby měření

konstrukce_a_vlastnosti_psychodiagnostickych_metod_image_0.wmf

objektivita (jak test používat)→reliabilita → validita → normalizace (ta by měla být našim cílem)

Konstantní chyby (ovlivňují validitu)

1. důsledkem nepřísti měření v psychologii (nějaký atribut měřeného objektu se porovnává s jiným atributem nástroje měření, který má pro daný měřící nástroj matematicky definovaný vztah s atributem měřeného objektu (např. Zavěšením závaží na pružinu, můžeme z jejího protažení odvodit váhu závaží – byť v psychologii moc o přesně definovaném matematickém vztahu moc mluvit nemůžeme – zasahují do toho právě ty chyby měření a proto bychom je prý jako psychologové měli znát)
2. v případě naprosté většiny metod si pak nemůžeme být jisti, zda tyto metody skutečně měří to, co o nich jejich autoři tvrdí
3. snaha eliminovat systematické chyby tedy souvisí s nejzákladnějším požadavkem kladeným na každou psychodiagnostickou metodu, totiž aby metoda skutečně měřila to co měřit má (např. u celé řady testů je nutné umět číst, ale bylo by chybou kdyby rozdíly ve výkonu v textu souvisely více se čtením než se zjišťovanou schopností)
4. validita se taky někdy definuje jako relativní nepřítomnot konstantních chyb měření v naměřených hodnotách

Osobní chyby (souvisí s objektivitou)

1. způsobovány konkrétním člověkem, který administruje, vyhodnocuje a interpretuje výsledky psychologického vyšetření
2. může zkreslit instrukce; nesprávně měřit čas; vytvořit atmosféru, která nepřispěje k žádoucímu výkonu probanda v testu; může špatně vypočítat dílčí skóry; ve fázi interpretace může dát výsledky do neadekvátního kontextu; při aplikaci si mohou dva různí odborníci po přečtení téže zprávy z vyšetření učinit dva různé závěry
3. pro omezení této chyby slouží –>

standardizace: v užším slova smyslu to spočívá ve stanovení takových pravidel administrace, skórování a interpretace výsledků měření, že jsu co nejpřesněji definovány podmínky, za kterých má měření probíhat.

1. Zajišťuje, aby všichni testovaní měli stejné podmínky
2. Záznamové archy, testové sešity, klíč pro vyhodnocení odpovědí, manuál,…
3. Instrukce jsou pokyny pro probandy, jak mají postupovat v psychodiagnostické situaci;
4. Hlavní požadavek kladený na instrukci spočívá v její jasnosti a srozumitelnosti pro všechny skupiny potenciálních probandů. Pokud se např. bude test používat pro vyšetření osob s nižším intelektem nebo vzdělání, měly by být formulace tak srozumitelné, aby jim i tyto osoby rozuměly.
5. Situace vyšetření – přiměřené osvětlení, teplota, hluk, striktní dodržení časového limitu
    1. Hrát roli může i pohlaví administrátora, popř. jeho/její fyzická atraktivita nebo rasová příslušnost

Proměnné chyby (ovlivňují reliabilitu)

1. O jejich existenci je možné se velmi jednoduše přesvědčit tak, že se provádí opakované měření téže veličiny. Jednotlivé naměřené hodnoty se od sebe liší, a to v závislosti na velikosti měřené hodnoty a přesnosti použité procedury měření
2. Na základě statistických analýz opakovaných měření je pak možné stanovit, jaký díl variability měření odpovídá skutečně existujícím rozdílům v měřených hodnotách a jaký díl variability měření odpovídá skutečně existujícím v měřených hodnotách a jaký díl takto vysvětlit nelze; ten je pak připisován na vrub náhodným vlivům
3. Pokud nám tedy často dává stejná metoda různé výsledky, není reliabilní (nelze se na i spolehnout, je příliš závislá na vnějších vlivech)

Interpretační chyby

- Chyby interpretace naměřených hodnot spočívají v nedodržení požadavku, aby se naměřená hodnota (tzv. hrubý skór) vždy interpretovala v nějakém kontextu

Řeší se tzv. normalizací - tedy vytvořením norem pro daný test

Reliabilita

1) reliabilita je charakteristika psychodiagnostické metody, která uvádí relativní nepřítomnost proměnných chyb v měření,

1. reliabilita je v podstatě jiný název pro spolehlivost nebo přesnost metody měření

Test-retest reliabilita (reliabilita jako stabilita v čase)

Zadáme jeden a ten samý test dvakrát v různém čase a korelujeme výsledky – jakékoliv rozdíly se považují za zdroj chyby

1. Tím pádem se předpokládá, že se měřený rys v čase nemění
2. Tento předpoklad může být nerealistický – nelze tedy používat např. pro zjišťování reliability metod pro zjišťování přechodných psychických stavů – např. dotazníků depresivity
3. po příliš krátké době oddělující obě testování může dojít ke vzniku jiného artefaktu (tzn. jevu, který vzniká nesprávným použitím metody) klamnému nadhodnocení reliability vlivem zapamatování odpovědí na jednotlivé otázky
4. doporučuje se odstup 3 měsíce

Reliabilita paralelních forem

1. Dva paralelní testy měří tentýž konstrukt stejným způsobem
2. Paralelní formy by měly mít stejné průměry hrubých skórů, stejné směrodatné odchylky a stejné korelace s jinými měřícími nástroji v jakékoli populaci
3. Pro zjištění reliability se oba testy korelují a srovnávají se průměry a rozptyly jednotlivých položek
4. Problémem je vytvořit skutečně paralelní formy

Split-half reliabilita

1. Odhad reliability získaný rozdělením testu na dvě poloviny
2. Řeší problémy předchozích dvou způsobů odhadu, ve kterých se možné reálné fluktuace rysu automaticky považují za chyby, přestože tomu tak být nemusí (tím by byla skutečná reliabilita testu podhodnocena)
3. Aby nebylo nutné složitě konstruovat paralelní formu testu, která nikdy nebude zcela dokonalá, administruje se pouze jeden test a pak se z důvodu analýzy rozdělí na dvě poloviny
4. Problém – reliabilita testu roste s počtem položek – tím pádem rozdělení na dvě poloviny snižuje odhad reliability
5. Nutná korekce pomocí Spearman-Brown prophecy formula - tzv. Spearman-Brownův prorocký vzorec:-) – prorocký, protože umožňuje odhadnout, nakolik zvýší reliabilitu, když přidáme určité množství položek

Kuder-Richardsonova reliabilita

1. Předchozí přístup se nehodí při zjišťování reliability výkonových testů s rychlostní složkou, protože je známo, že stupňující se obtížnost položek u testů s rychlostní složkou vede k tomu, že subskóry vytvořené z několika po sobě následujících položek spolu obvykle korelují nízce, přestože měří stejnou schopnost
2. Kuder-Richardsonův přístup pracuje s tzv. *homogenitou*
3. *Homogenita* - konsistence výkonů v sadě položek
4. V ideálním případě jsou položky testu s rychlostní složkou seřazeny podle vzrůstající obtížnosti. Dá se tedy očekávat, že osoba bude nějakou dobu řešit položky bez obtíží, a to až do té chvíle, kdy dosáhne svého osobního „stropu„, tzn. položky, za kterou už nevyřeší žádnou další. V ideálním případě by pak výkon všech osob vypadal tak, že by dosáhly bodu, před kterým vyřešily všechny položky a za kterým nevyřešily žádnou (samozřejmě každá osoba by tento bod měla jinde, podle své úrovně měřené schopnosti). V takovém případě by se test vyznačoval dokonalou reliabilitou podle Kudera a Richardsona
5. stejně důležité je, aby test měřil pouze jeden rys nekontaminovaný žádným jiným - v takovém případě se o testu říká, že není tzv. jednorozměrný a požadavek homogenity bývá obvykle nesplnitelný
6. odhad reliability se provádí pomocí koeficientu KR20 – použitelný pouze u dichotomických položek (správně/špatně)

Reliabilita podle Hoyta

1. Je podobný teorii zobecnitelnosti
2. Hoyt (1941; In: Brzeziński, 1999) tvrdí, že kolísání výkonu osoby od položky k položce by se nemělo považovat za chybu, ale za reálně existující intraindividuální rozdíly, které by tudíž neměly zkreslovat (snižovat) odhad reliability
3. Rozeznává tři zdroje variability, tzv. komponenty rozptylu, jejichž relativní důležitost lze při správně provedeném výzkumu vzájemně porovnat pomocí analýzy rozptylu

a) interindividuální rozdíly, tzn. skutečné rozdíly mezi lidmi, kvůli kterým se vlastně měření v psychologii a psychodiagnostika vůbec provádí

b) intraindividuální rozdíly, tzn. skutečné kolísání měřených hodnot

1. chyby měření (variabilita chyb)

Faktory ovlivňující odhad reliability

Metoda odhadu reliability

1. rozhodnutí o volbě argumentů pro reliabilitu meotdy by mělo být založeno ne na pragmatické snaze o získání co nejlepších parametrů, ale spíše na vlastnostech diagnostického nástroje, které samy vhodnou metodu indikují (přítomnost eychlostního

faktoru; typ položek, z nichž jsou tvořeny škály; účely metody). V zásadě vša platí, že nemá smysl volit metody tak, aby se navzájem duplikovali,
2. paralelní formy dávají nejnižší odhad, split-half nejvyšší (pozn. u reliability vždy mluvíme o odhadu)

Délka testu

1. čím více položek máme (teoreticky je totiž položek zjišťujících daný atribut nekonečně mnoho), tím více je náš test reprezentativní pro daný rys (za předpokladu, že máme kvalitní položky) – tím pádem máme i vyšší reliabilitu

Heterogenita skupiny

1. variabilita měřeného rysu ve výběrovém souboru – při příliš homogenním souboru se snižuje variabilita pravých skórů, přičemž variabilita chyb měření zůstává stejná – to znamená nižší odhad reliability
2. kde není variabilita, tam není realabilita

časové aspekty

1. u rychlostních testů je nadhodnocena reliabilita ve smyslu vnitřní konzistence
2. možná je to tím, že lidé vykazují větší konzistenci v tempu práce, než v kvalitě výkonu
3. Cronbach píše, že pracovní tempo je samo o sobě rysem respondenta. To by znamenalo, že výkon v každém rychlostním testu je dán dvěma latentními proměnnými – cíleně měřenou charakteristikou a rychlsotní složkou, které jsou navzájem pravděpodobně nekorelované
4. z uvedených důvodů je užitečné vědět, do jaké míry jsou výkony v určitém testu ovlivněny časovým omezením, a tudíž rychlostí, jakou jsou probandi schopni pracovat.

Konstrukce a administrace metody

1. špatná standardizace zadání testu výrazně snižuje reliabilitu
2. Vlastnosti testu samého: ne – vzájemně závislé položky (nevím jednu, nemohu odpovědět na druhou); ne téměř identické položky; ne chytáky; taky emocinálně zbarvené položky berou do hry další atribut
3. Forma položky
4. Vyšetřované osoby
5. Administrátor

Validita

K validitě se dá přistupovat ze tří stran

1. z hlediska obsahu metody a situace testování a chování testované osoby v této situaci (těmto zdrojům důkazů o validitě se říká obsahové zdroje)
2. z hlediska vztahu mezi skórem testu a skórem nějakého vnějšího kritéria (empirické zdroje)
3. z hlediska začlenění měřeného atributu do vztahů dalších atributů v rámci teorie, ve které se měřený atribut definuje (konstruktové zdroje)

Obsahové zdroje důkazů o validitě

1. míra, do jaké je obsah psychodiagnostické metody, situace testování a chování požadované od probanda, který se v této situaci nachází, ve shodě s účelem metody

Zjevná validita metody (face validity)

1. zjevnou validitou se může pyšnit metoda měření, u které proband, který se stane jejím objektem měření, dokáže víceméně trefně odhadnout, co se metoda pokouší změřit
2. Zdánlivá validita může zvýšit motivaci testovaných osob – mají pocit, že daný test má smysl, a tak mu věnují více energie – např. Rorschach nemá pro většinu lidí vůbec žádnou zdánlivou validitu a z toho se může odvíjeti i jejich motivace test dělat
3. U výkonových testů není problémem, u osobnostních a jiných problémem být může
4. může být i zdrojem zkreslení – ldyž na výsledcích testování závisí další osud probanda (někdy tak může být lepší, když test zjevnou validitu nemá)

Výběrová validita

1. posouzení obsahové validity spočívá v oslovení nějakého počtu osob, které lze považovat za odborníky v daném oboru a od nichž chceme expertní posudek
2. spočívá v posouzení adekvátnosti výběru obsahu položek vzhledem k rysu, který má být metodou měřen
3. používá se zejména u výkonnových testů nebo testů schopností a vědomostí

Faktorová validita

-Obsah metody se posuzuje na základě faktorové analýzy dat představujících položky testu řešené nebo zodpovězené dostatečně početným souborem osob vybraných z relevantní populace.

Empirické důkazy o validitě

1. jako argument svědčící o tom, že metoda skutečně měří atribut, pro jehož měření byla vytvořena, samotné důkazy o obsahové validitě nestačí. Mohou prokázat pouze to, že byly podniknuty adekvátní pokusy o pokrytí dané obsahové domény.
2. Pro doložení validity obecně je nutné nějaké kritérium měřeného rysu, které v případě, že je metoda validní, má vztah se skórem získaným touto metodou
3. Největší problém empirické validity představuje nalezení vhodného kritéria
4. Důležitá je reliabilita kritéria a také heterogenita výběrového souboru

Prediktivní validita metody

Týká se míry, do jaké je hodnocený test použitelný pro předpovídání budoucích hodnot nějakého krtéria např. Nakolik je naměřená hodnota IQ vhodným prediktorem akademického úspěchu

Souběžná validita metody

1. souběžná validita dvou a více metod
2. Náš test se koreluje s výsledky testu měřícího stejný rys administrovaný ve stejnou dobu

Inkrementální validita metody

1. Inkrementální validitou se může chlubit takový test, který přináší vzhledem k nějakému účelu informaci, kterou nelze získat z jiných zdrojů
2. Tzn. inkrementální validita určuje, nakolik nám určitý test dopomáhá k našemu cíli – např. zpřesnění diagnózy, určení vhodného povolání atp.
3. Jako příklad lze použít inteligenční test, který obsahuje různé subtesty pro měření dílčích složek inteligence. Tyto subtesty mají vůči sobě inkrementální validitu v situaci, kdy je cílem zjistit nejen celkovou úroveň inteligence, ale také strukturu schopností

Diferenciální validita metody

1. Je určena tím, že výsledky našeho testu **nekorelují** s těmi, se kterými korelovat nemají

Konstruktové důkazy o validitě

1. přišli s tím Cronbach a Meehl (1955), když vyslovili požadavek, aby každá validizační studie obsahovala také hodnocení měřeného atributu jako konstruktu v rámci relevantní psychologické teorie
2. teprve v kontextu dané teorie je podle jejich názoru možné komplexně posousdit všechny obsahové aspekty, které by měla mít metoda měřící daný atribut, a také všechby hypoteticky důležité vztahy měřeného atributu s jinými atributy.
3. Cronbach a Meehl (1955) navrhli také logický postup, který by se měl při posuzování konstruktové validity použít:

1) na začátku se přijme předpoklad, že daný test měří rys A

2) posoudí se současný stav teorie týkající se měřeného rysu A

3) zformulují se hypotézy, týkající se vztahů rysu A s ostatními rysy

4) hypotézy se empiricky ověří

1. Tím pádem zjišťování konstruktové validity v sobě obsahuje všechny výše jmenované aspekty hodnocení validity

Helmstadter (1964) uvádí některé důležité aspekty související s procedurami hodnocení konstruktové validity.

1) Někdy se zjistí, že kritérium používané v raných stadiích vývoje testu je méně validní než test. To je již zmíněný případ testů inteligence, které v současné době považujeme za přesnější než hodnocení učitelů. Dá se říci, že se jedná o ideál, ke kterému se v rámci psychometriky snažíme dospět.

2) Uživatel testu potřebuje znát i teorii a důkazy její platnosti. Pokud není teorie pochopena, není možné adekvátně interpretovat výsledky testování

3) Konstruktovou validitu testu lze hodnotit pouze tehdy, byl-li test zveřejněn a zpřístupněn psychologické veřejnosti. Zejména je na miste opatrnost při používání testů, jejichž konstrukce a postup skórovaní jsou tajné

4) Každý výsledek ve shodě s teorií postupně podporuje validitu, ale dobře metodologicky a teoreticky fundovaný negativní nález muze vše najednou vyvrátit

5) Otázka po validitě testu je v konečném důsledku naivní, protože test nelze nikdy zcela validizovat (Cronbach a Meehl, 1955) a test může byt validní pouze vzhledem k nějakému účelu.

Standardizace

1. je nutnou podmínkou objektivity

2. Zajišťuje, aby všichni testovaní měli stejné podmínky
3. Záznamové archy, testové sešity, klíč pro vyhodnocení odpovědí, manuál,…
4. Instrukce jsou pokyny pro probandy, jak mají postupovat v psychodiagnostické situaci;
5. Hlavní požadavek kladený na instrukci spočívá v její jasnosti a srozumitelnosti pro všechny skupiny potenciálních probandů. Pokud se např. bude test používat pro vyšetření osob s nižším intelektem nebo vzdělání, měly by být formulace tak srozumitelné, aby jim i tyto osoby rozuměly.
6. Situace vyšetření – přiměřené osvětlení, teplota, hluk, striktní dodržení časového limitu
7. Hrát roli může i pohlaví administrátora, popř. jeho/její fyzická atraktivita nebo rasová příslušnost

Normalizace

1. Týká se tvorby norem, které pak slouží k porovnání výsledků jednotlivce s výsledky nějaké vhodně definované populace
2. normy dávají testovým skórům psychologický smysl tím, že umožňují jejich interpretaci
3. Normalizací rozumíme srovnání individuálních výsledků s normami, získanými vyšetřením velkého reprezentativního vzorku osob (norma je chápána jako průměrný výkon nebo typickou reakci příslušného vzorku populace). Pro snazší srovnání s normou se výkony dosažené v testu, tzv. **hrubé skóry**, převádějí na **vážené** nebo **standardní skóry.** Druhy norem:
4. **věková norma** – příkladem je **deviační inteligenční kvocient,** kt. udává, jak mnoho se respondent liší od věkově příslušného populačního průměru
5. **kvantilová norma** – typickým příkladem jsou **percentily**, kt. udávají kolik procent populace je horší v příslušném kritériu než vyšetřovaná osoba - osoba s hodnotou v testu 95 percentil = její výkon je lepší než výkon 95% populace nebo **decily,** ty jsou desetkrát větší → 1 decil = 10 percentilů
6. **stenová = standardní norma** – způsob uvádění normalizovaných výsledků v hodnotách ležících v intervalu 1-10