

# Přednáška 8: Férovost a zkreslení při testování

---

24. 11. 2020 | PSYn4790 | Psychometrika: Měření v psychologii  
Katedra psychologie, Fakulta sociálních studií MU

Hynek Cígler | [hynek.cigler@mail.muni.cz](mailto:hynek.cigler@mail.muni.cz)

Co si představíte  
pod termínem **férovost**...  
... v psychodiagnostice?  
... a v psychometrice?

# Férovost v psychometrice

---

2. kapitola českého překladu *Standardů pro pedagogické a psychologické testování* (AERA, 2001).

- Doporučuji vydání 2014 v Aj
- A to i pro studium PSYn4020/PSYn5340.

„Férovost“ a s ní související téma multikulturního testování je jedním z důležitých témat současné psychometriky (zejména v rámci tzv. edukativního testování).

# Klíčové pojmy

---

## **Přístupnost** (accessibility).

- Měřený rys je *stejně dostupný* u všech potenciálních probandů.
- Příslušnost ke skupině probandů neovlivňuje výsledek v testu *po kontrole rysu*.
- Např. zrakové/sluchové znevýhodnění, znalost jazyka apod.

## **Univerzální design.**

- Charakteristika testu, která zajišťuje přístupnost.
- Např. snaha o vyřazení položek se silnou kulturní specificitou.
- Nebo zvážení rozdílnost účelu testu napříč skupinami probandů.

# 4 základní významy férovosti (AERA, 2014)

---

Férovost zacházení během testování.

- **Psychodiagnostika.**
- „Objektivita“, rovné zacházení... se všemi zacházím stejně.
- *Standardizace I* dle Urbánka ([2010](#)).

Nepřítomnost testového zkreslení.

- **Psychometrika.**
- „Test bias“, „item-bias“.
- Test a položky měří u všech stejný rys.
- DIF, DTF, DPF, invariance atd.

Férovost jako přístupnost, otevřenost.

- **Psychometrika, psychodiagnostika, teorie.**
- „Accessibility“, „provability“.
- Vlastnost je u respondenta měřitelná.

Férovost jako interpretace individuálního skóre pro daný účel

- **Psychodiagnostika.**
- Zvážení jedinečnosti každého respondenta.
- Jaká individuální specifika ovlivňují výkon.
- Akomodace, individuální úpravy testu.
- *Důsledky testování* (Messick).

# Bias = zkreslení

---

## **Bias = systematické zkreslení** testových výsledků

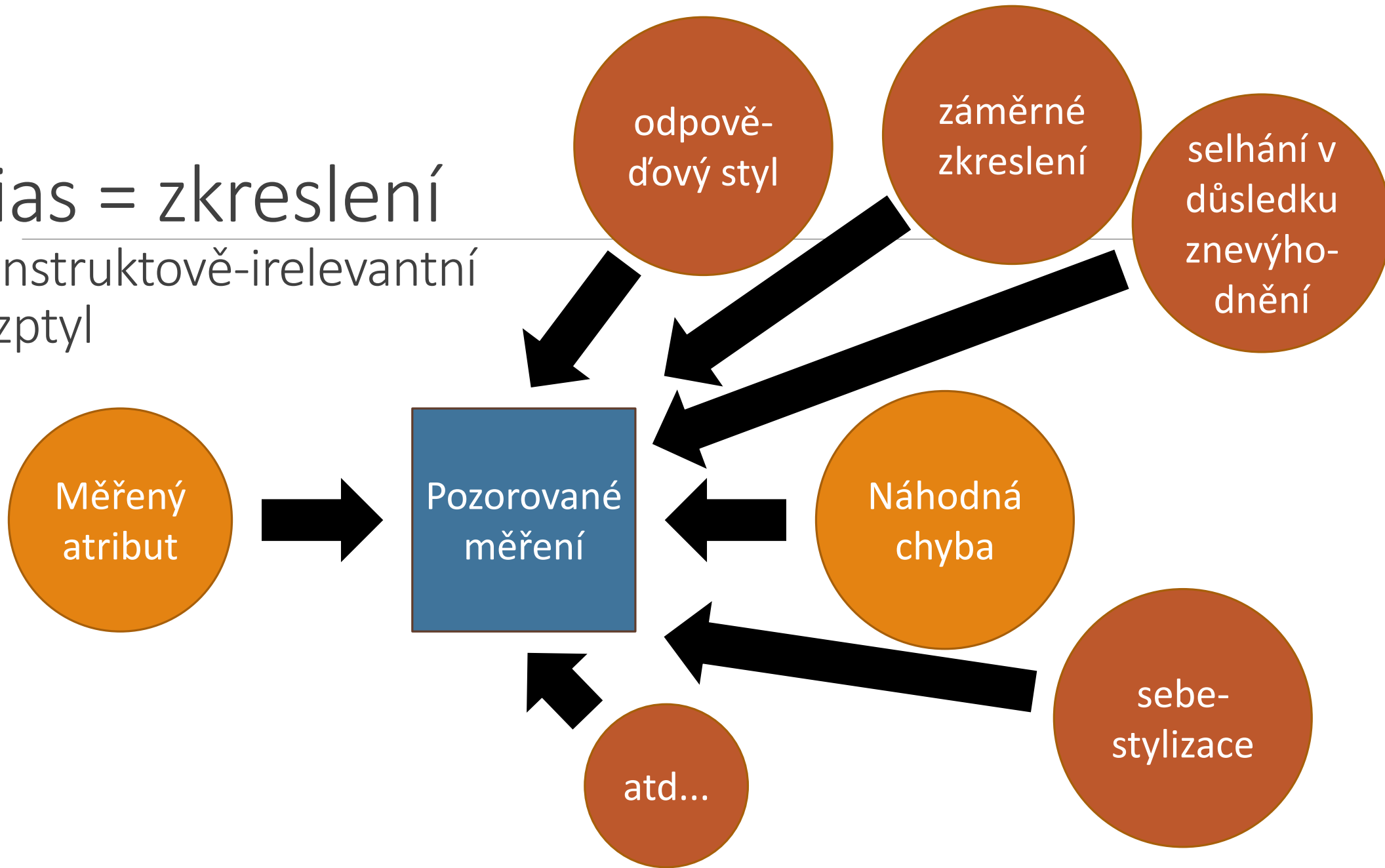
- Reliabilita: náhodné chyby měření, není tedy otázkou „bias“.
- Úvaha o zkreslení patří do validity, ale z praktických důvodů je vyčleňována.
- Slovem zkreslení označujeme nenáhodné, specifické chyby měření.
- *„Měří test jinak pro některé populace než pro jiné?“*
- *„Měří test jinak pro některé specifické osoby?“*
- *„Měří test obecně spravedlivě?“*

## Může znamenat, že v různých populacích např.:

- Je test/položka příliš snadná/obtížná.
- Má test/položka jiný vztah k rysu.
- Test má jinou faktorovou strukturu.
- Test měří zcela či částečně něco jiného.
- ...
- *Výkon v testu je ovlivněn systematicky něčím, co nemá souvislost s tím, co chci měřit („konstruktově irelevantní rozptyl“).*

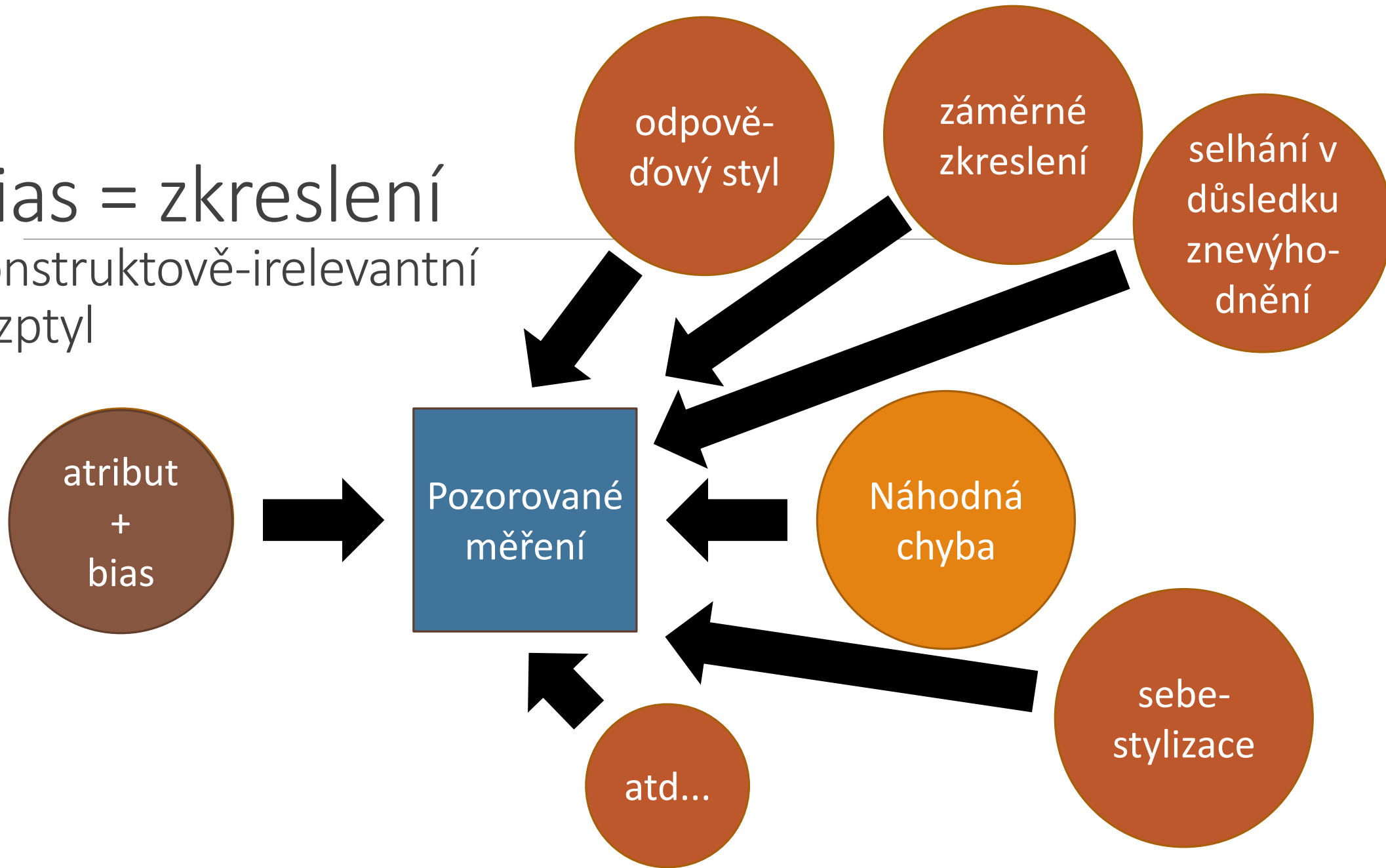
# Bias = zkreslení

Konstruktově-irelevantní rozptyl



# Bias = zkreslení

Konstruktově-irelevantní rozptyl





# Důsledky zkreslení

---

## PŘEDPOKLAD

CTT model měření pro nezkreslený pozorovaný skór  $X$ :

$$X = T + e$$

Reliabilita:

$$r_{xx'} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_e^2}$$

Protože  $\sigma_T^2 < \sigma_T^2 + \sigma_{B_1}^2 + \sigma_{B_2}^2 + \dots + \sigma_{B_n}^2$ , pak zpravidla platí  $r_{yy'} > r_{xx'}$ .

- Jinými slovy, testové zkreslení velmi často **zdánlivě zvyšuje reliabilitu**.

Protože ale  $B_1 + B_2 + \dots + B_n$  nesouvisí s případným kritériem, snižuje validitu.

## DŮSLEDEK ZKRESLENÍ

CTT model měření pro zkreslený pozorovaný skór  $Y$ :

$$Y = T + B_1 + B_2 + \dots + B_n + e$$

Reliabilita:

$$r_{yy'} = \frac{\sigma_T^2 + \sigma_{B_1}^2 + \sigma_{B_2}^2 + \dots + \sigma_{B_n}^2}{\sigma_T^2 + \sigma_{B_1}^2 + \sigma_{B_2}^2 + \dots + \sigma_{B_n}^2 + \sigma_e^2}$$

# Příklad potenciálních oblastí zkreslení

---

**Objektivita:** Zkreslení na úrovni examinátora a testové situace, nestranné zacházení.

- Je zacházeno se všemi respondenty stejně?

**Response bias:** Zkreslení odpovědí na úrovni respondenta.

- Záměrné i nezáměrné zkreslení odpovědi respondentem na úrovni položky.

**Item bias:** Systematické zkreslení položky.

- Systematické rozdíly mezi osobami/skupinami v odpovědi na položku, nevysvětlitelné úrovní rysů.

**Test bias:** Systematické zkreslení testu.

- Týká se systematického zkreslení celkových skóre/výsledků testu napříč skupinami respondentů.

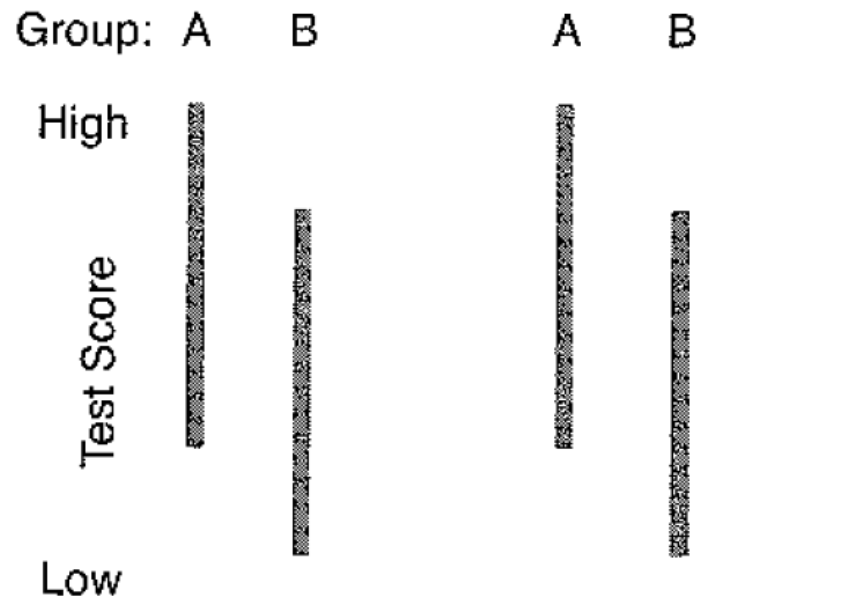
**Predictive bias:** Systematické zkreslení testu.

- Rozdílný vztah testových výsledků s kritériem.

# Test bias, test fairness

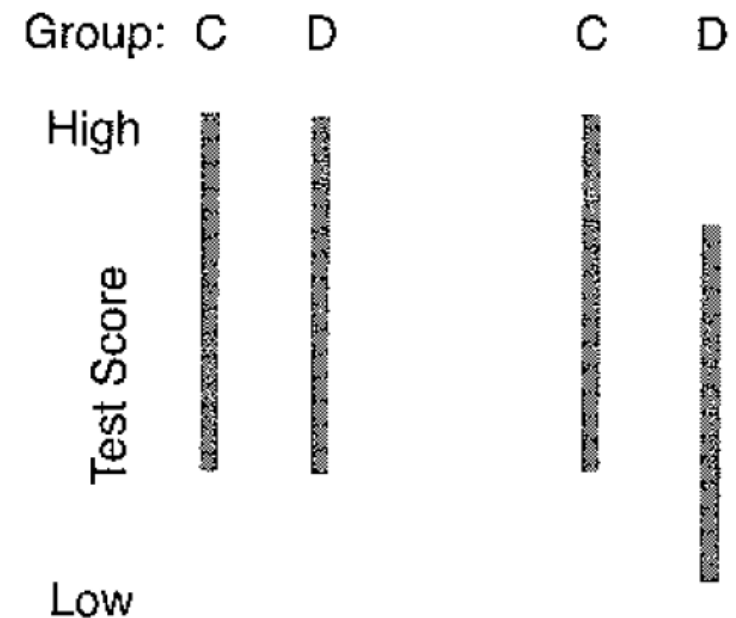
## A Fair Test, Lack of Bias

Real Status on the Trait or Ability      Performance on the Test



## A Biased, Unfair Test

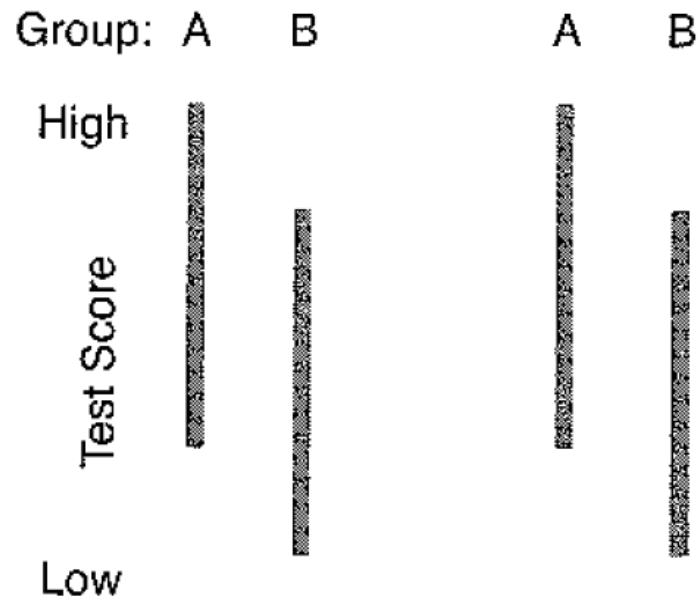
Real Status on the Trait or Ability      Performance on the Test



# Test bias, test fairness

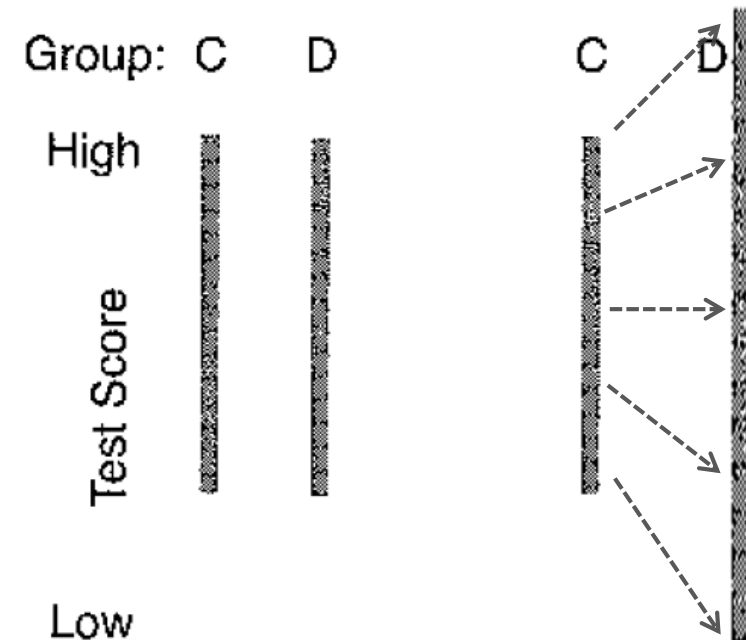
## A Fair Test, Lack of Bias

Real Status on the Trait or Ability      Performance on the Test



## A Biased, Unfair Test

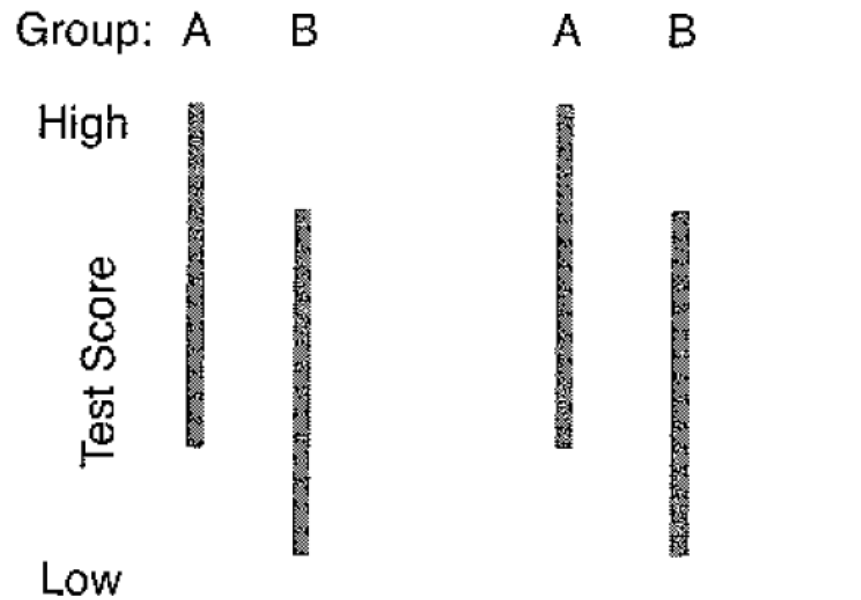
Real Status on the Trait or Ability      Performance on the Test



# Test bias, test fairness

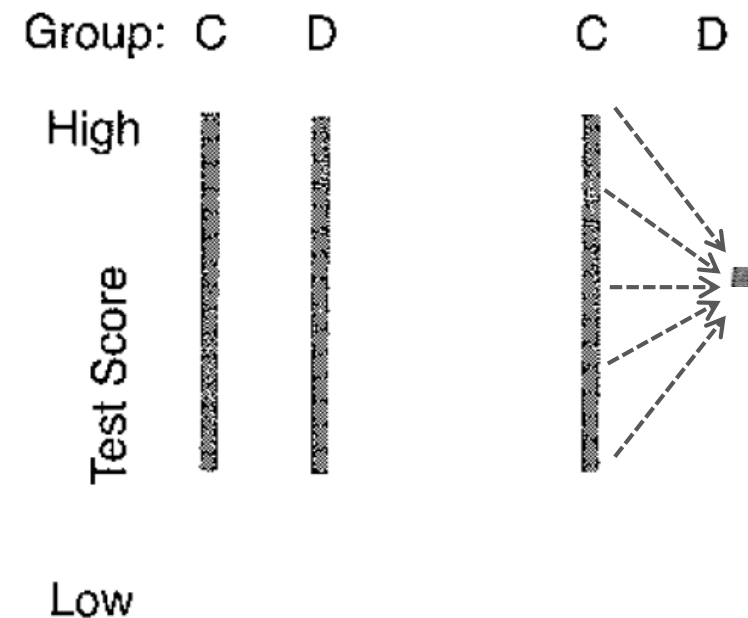
## A Fair Test, Lack of Bias

Real Status on the Trait or Ability      Performance on the Test



## A Biased, Unfair Test

Real Status on the Trait or Ability      Performance on the Test



# Zdroje ohrožení férovosti testování

---

## **Obsah testu**

- Který znevýhodňuje některé skupiny, osoby atd.

## **Kontext testové situace**

## **Odpovědi na položky**

- Formát položek, interpretace při kvalitativním skórování výsledků...

## **Příležitost k přípravě na test**

A další...

# Kontext testové situace

---

Tohle je otázka spíše do psychologické diagnostiky/etiky.

Cílem je zajistit, aby každý respondent měl možnost projevit ty stejné schopnosti ve stejné míře.

- APA standard 7.12: *„Testování nebo hodnocení by mělo probíhat takovým způsobem, aby se všem testovaným osobám dostalo stejného nebo srovnatelného zacházení během všech fází testování.“*

Administrátor testu rovněž musí být kompetentní s konkrétním testem pracovat (školení, zácvik...).

# Možnost přípravy

---

Všichni respondenti musí mít shodné možnosti zácvičku, poučení o cíli testování...

- Na tohle pozor! Běžná praxe neomlouvá...

Př. 1: „Tajné“ informace o způsobu dopravně-psychologického vyšetření.

Př. 2: Placené (a drahé) přípravné testy na přijímačky.

Př. 3: Neformálně dostupné informace o průběhu forenzního vyšetření.

Př. 4: Různý způsob informování před zahájením vyšetření.



# Férovost jako přístupnost

---

## **Příklad: Přijímačky do bc studia na FSS**

2 testy: studijní předpoklady (váha 0,4), ZSV (váha 0,6)

Studijní předpoklady – na výběr:

- SCIO (až 5 pokusů, bere se nejlepší)
- TSP od MU (1 pokus)

ZSV – jediná možnost:

- SCIO (až 5 pokusů, bere se nejlepší)

Jaké jsou nevýhody daného designu z hlediska psychometrie?

Co byste studentům řekli, aby měli rovné podmínky?

Simulace:

<http://fssvm6.fss.muni.cz/prijimZk/>

# Response bias

---

Jde o určitý styl odpovídání specifický konkrétnímu respondentovi v konkrétní situaci, který znehodnotí/zneplatní testové výsledky (sníží jejich validitu).

Nahodilé odpovědi a záměrné zneplatnění výsledků.

Zkreslení v užším významu (práce Paulhuse a kol.<sup>1</sup>).

- Simulování a sebeznevýhodňování (záměrné). Tzv. „impression management“.
- Sociální žádoucnost a nezáměrné zkreslení. Tzv. „self-deception“.

Response style

- Tendence k souhlasu nebo nesouhlasu.
- Tendence k extrémním nebo průměrným odpovědím.

Hádání, tipování.

<sup>1</sup> Řada dílčích publikací o self-presentation, overclaiming, self-management atd.

# Response bias – řešení

---

Změna setingu testové situace, aby respondent nebyl motivován výsledky zkreslovat.

- Anonymita, redukce stresu, srovnání úrovně motivace...

## Úprava položek

- Jednoduché položky – krátké jednoznačné stimuly, krátké jednoznačné a „ne-extrémní“ distraktory.
- U delších odpověďových (Likertových škál) je větší prostor pro zkreslení.
- Zajištění absence chybějících odpovědí.
- Rozdílná valence položek (negativní skórování).
- Nucená volba (pak ale obtíže s psychometrickým zpracováním).

## Odhalení zkreslení

- Tzv. „validizační škály“ či „lži škály“ (např. v případě MMPI-II 6 různých škál).
- Dodatečné testy (Malingering scale – máme v KDM).
- Netestová detekce 😊.

# Metody ověření systematického zkreslení

---

**Expertní panelová review:** Obsahová validita.

**Diferenciální fungování položek:** Vnitřní struktura testu.

**Testová invariance:** Vnitřní struktura testu.

**Diferenciální predikce testu:** Prediktivní/kriteriální validita

# Panel review

---

Zejména v případě high-stakes testů.

Pečlivá volba tzv. expertního panelu (Subject Matter Experts, SME).

SME panel vytváří, reviduje a připomínkuje položky a složení testu (zejm. didaktické a edukativní testy).

- Experti musí být experty na měřený konstrukt.
- Zároveň by ale měli dobře reprezentovat testovanou populaci.
- Muži i ženy, minority...
- Jsou ale SME z určité minority dobrými reprezentanty této minorit?

# Test bias, item bias: Férovost z hlediska psychometriky.

---

DIFFERENTIAL ITEM FUNCTIONING (DIF)

DIFFERENTIAL TEST FUNCTIONING (DTF)



# Test/item bias

---

Nelze odvodit bez dat (jen odhadovat).

- Empirické důkazy a technická řešení.

Respondent se snaží odpovídat pravdivě, ale test měří v různých skupinách něco jiného.

WAIS-III: *„Co uděláte, když najdete na zemi zalepenou poštovní obálku s napsanou adresou, známkou, ale bez razítka?“*

WISC-III: *„Co uděláte, když chcete uvařit čaj?“*

Skupiny: etnikum, pohlaví, jazyk, socio-ekonomický status, region...



# Dva hlavní empirické přístupy k férovosti

---

## **Na úrovni položky** (item bias analysis).

- Které položky (a zda ta která položka) vykazují rozdílný styl odpovídání napříč skupinami, který nelze přičíst rozdílům v úrovni latentního rysu?
- DIF analýza (differential item functioning)

## **Na úrovni testu** (test bias analysis).

- Do jaké míry test jako celek (soubor mnoha různých položek) měří ten stejný rys pro různé skupiny?
- Lze srovnávat naměřené skóry napříč skupinami?
- Invariance testu.

# Logika ověření zkreslení

---

## **Předpoklad férovosti:**

Atribut (latentní rys) „způsobuje“ pozorované odpovědi.

## **Systematické zkreslení znamená:**

Příslušnost ke skupině moderuje tento vztah.

- Zvyšuje/snižuje intercept závislé proměnné.
- Zvyšuje/snižuje regresní koeficient.
- Zvyšuje/snižuje reziduální rozptyl.

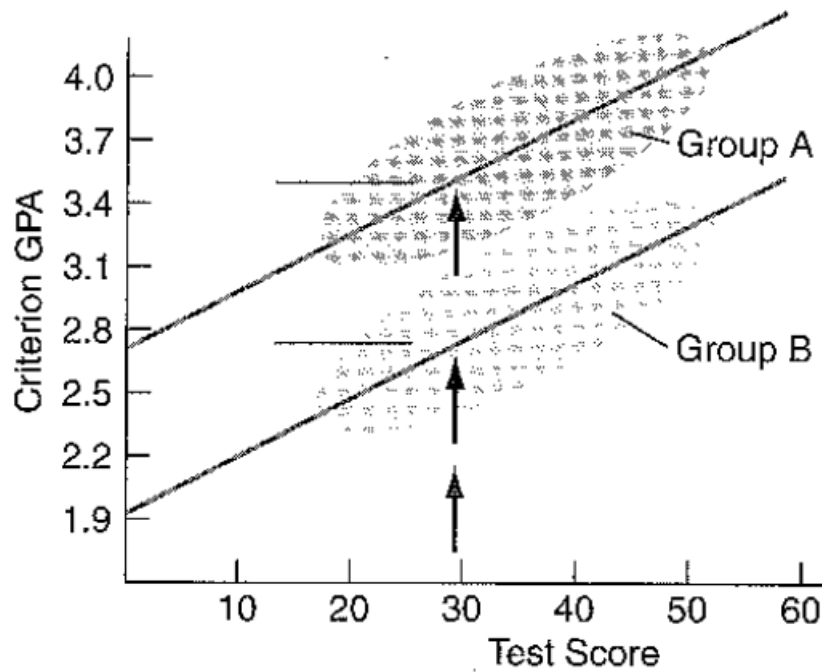
**Bias/zkreslení = moderace.**

# Test bias (zkreslení predikce)

např. přijímací zkoušky vs. státnice

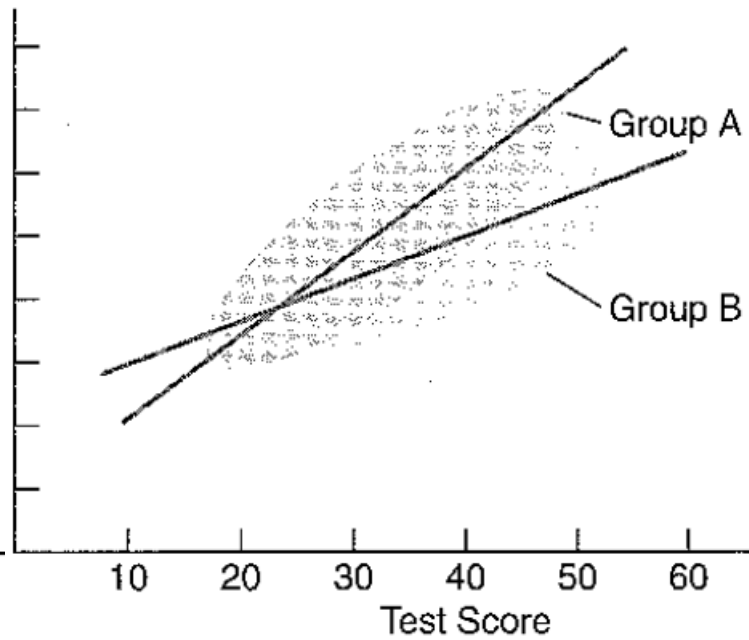
**situace A**

rozdíl v průměru predikce



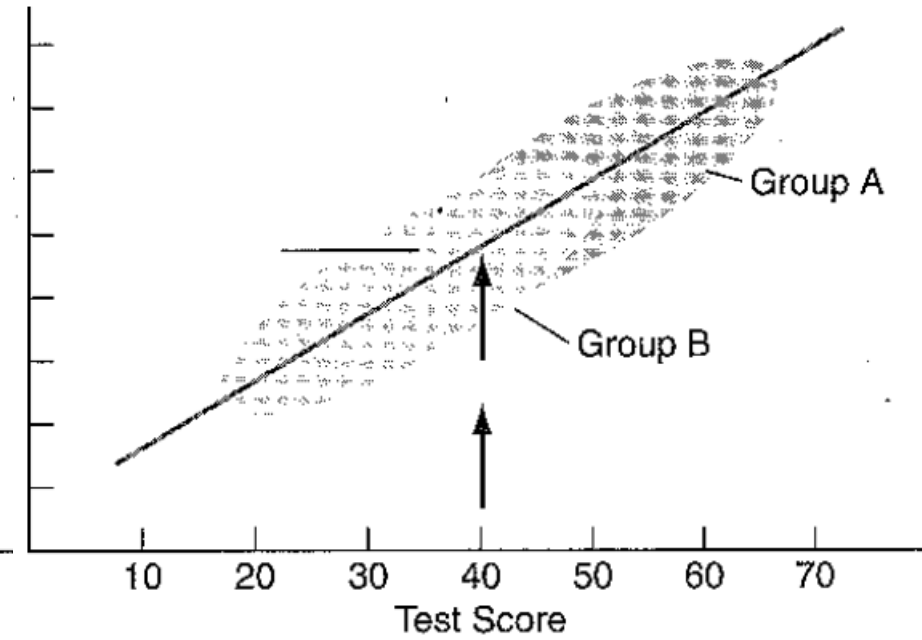
**situace B**

rozdíl v přesnosti (a průměru) predikce



**situace C**

férový test



# Test bias

---

Ověření typicky pomocí moderačního modelu (lineární i logistická regrese).

**Krok 1:** vytvoření interakční proměnné součinem prediktoru a moderátoru.

**Krok 2:** prostá lineární regrese

- Prediktivní nebo kriteriální validita.

$$Y = aX + b$$

- $Y$  – kritérium,  $X$  – výsledek testu
- $a$  – směrnice,  $b$  – průsečík.

**Krok 3:** přidání moderátoru do regrese.

- (Výhodnější je přidávat členy postupně.)

$$Y = aX + b + (cM + d(M \cdot X))$$

- $M$  – moderátor (skupina osob...)

Signifikantní F-test rozdílu 1. a 2. modelu ( $\Delta R^2$ ) → přítomnost zkreslení.

- sig.  $c$  → rozdíl v průměru predikce.
- sig.  $d$  → rozdíl v přesnosti predikce.

**Srovnáváme nestandardizované koeficienty!**

- Standardizované jsou ovlivněné populačními charakteristikami, které se lišit mohou.

# DIF v CTT

---

DIF = Differential Item Functioning

Nepružné, protože CTT a FA nedobře modeluje odpovědi na položku v závislosti na HS (předpoklad linearity odpovědí).

**Komparace ULI indexů:** Rozdělíme vzorek pro výpočet ULI napříč skupinami.

- ULI následně spočítáme pro celý vzorek, pro jednu i druhou skupinu.
- Jsou stejné? Jaká je korelace ULI napříč skupinami?

**Komparace popularit položek.**

- Je pořadí položek dle obtížnosti stejné napříč skupinami?
- Korelují popularity položek napříč skupinami?
- Spearmanova korelace obtížností položek.

# DIF v CTT (korektně)

---

**Mantelův-Haenszelův test.** Chí-kvadrát pozorovaných odpovědí pro každou úroveň HS a následná agregace výsledků.

Postupy založené na **logistické regresi**.

- Podobný přístup jako v případě test bias.
- Prediktorem je hrubé skóre, závislou odpověď na položku, moderátorem příslušnost ke skupině.
- binární položky: logistická regrese
- ordinální položky: ordinální nebo multinomická logistická regrese.

# IRT: Differential Item Functioning (DIF)

---

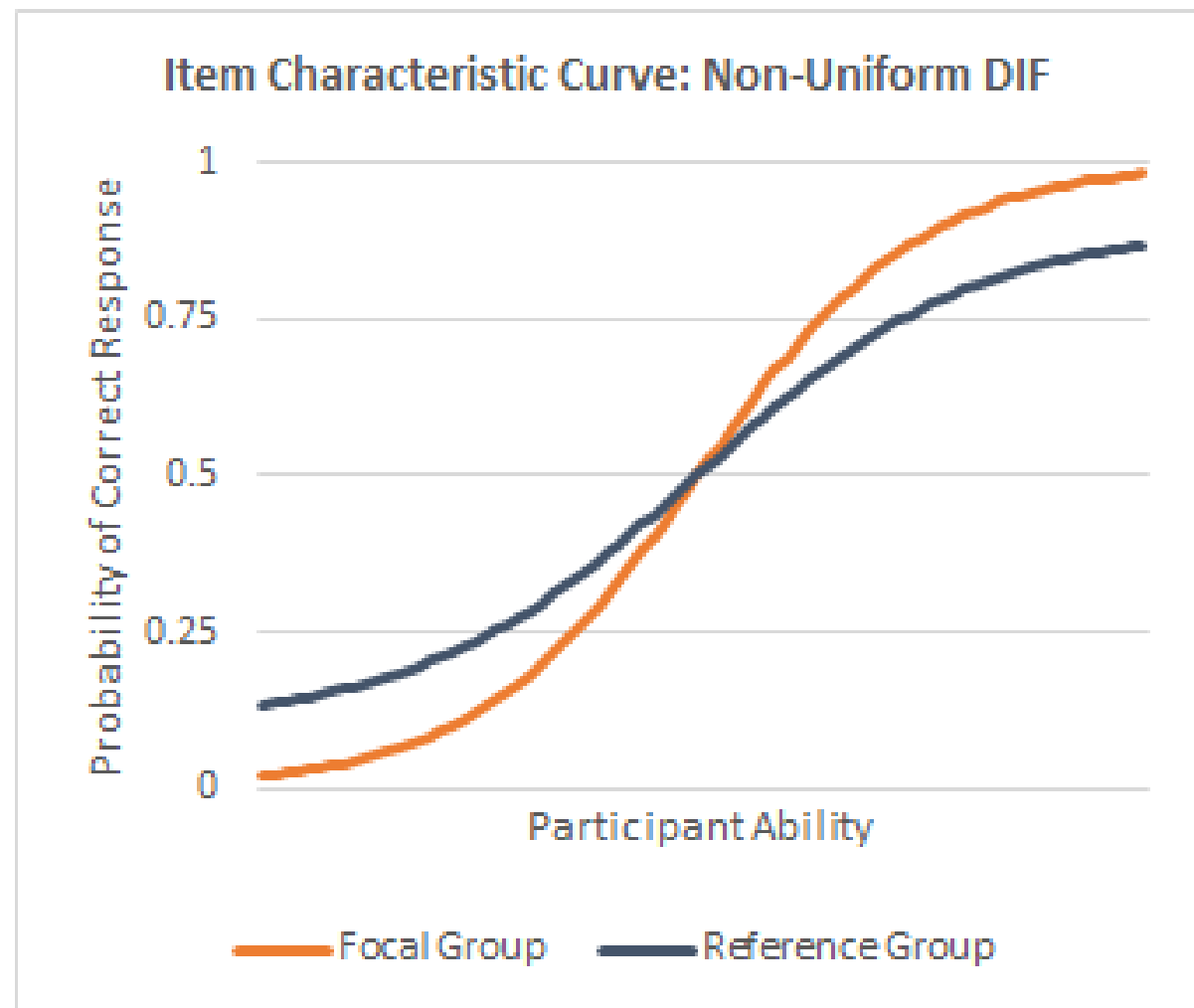
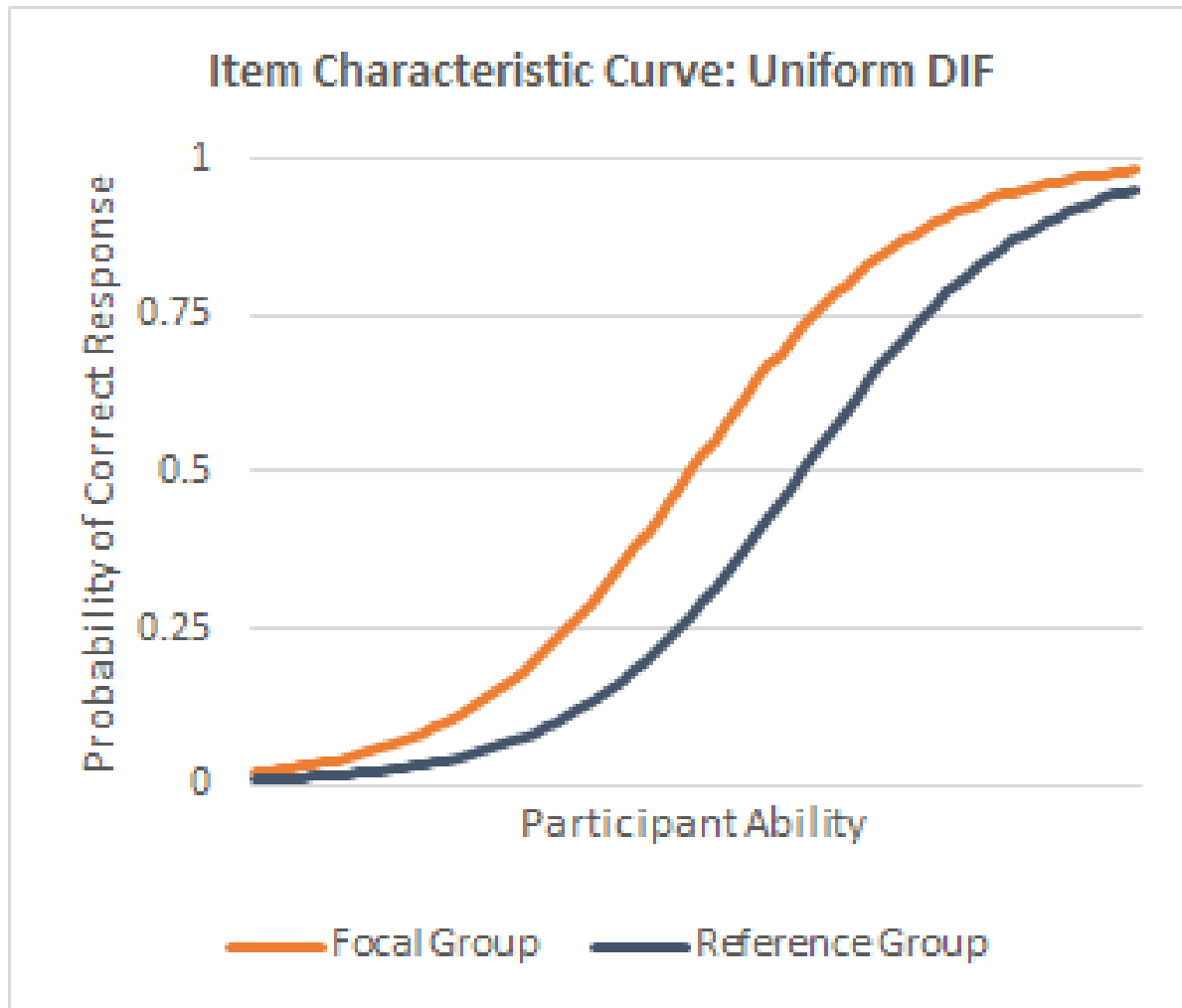
DIF analýza se používá se zejména v rámci IRT.

Obecný framework pro usuzování na neférovost jednotlivých položek.

- Některé postupy aplikovatelné i v CTT, ale IRT je výrazně vhodnější.
- V CTT je např. problematické testovat non-uniform DIF (viz dále), nebo DIF mezi skupinami, které se výrazně liší svým výkonem.

Základní princip:

**srovnání se charakteristická funkce položky napříč skupinami.**





# Příklad: Žádné DIF

Dotazník výšky:

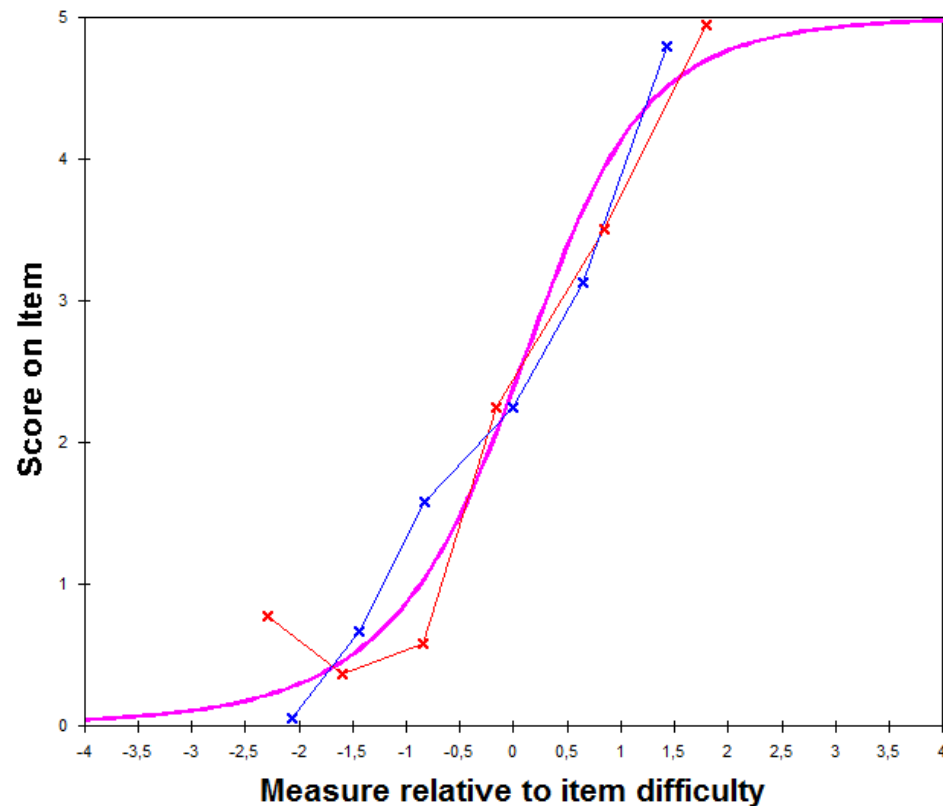
*Někdy se uhodím do hlavy o nízký strop, futro a podobně.*

DIF:

- t-test:  $t(86) = -0,31$ ,  $p=0,756$
- M-H:  $\chi^2(1)=0,44$ ,  $p=0,508$ .

Modrá muži, červená ženy.

4. Někdy se uhodím do hlavy o nízký strop, futro a podobně (DIF=\$S1W1)



# Příklad: Uniform DIF

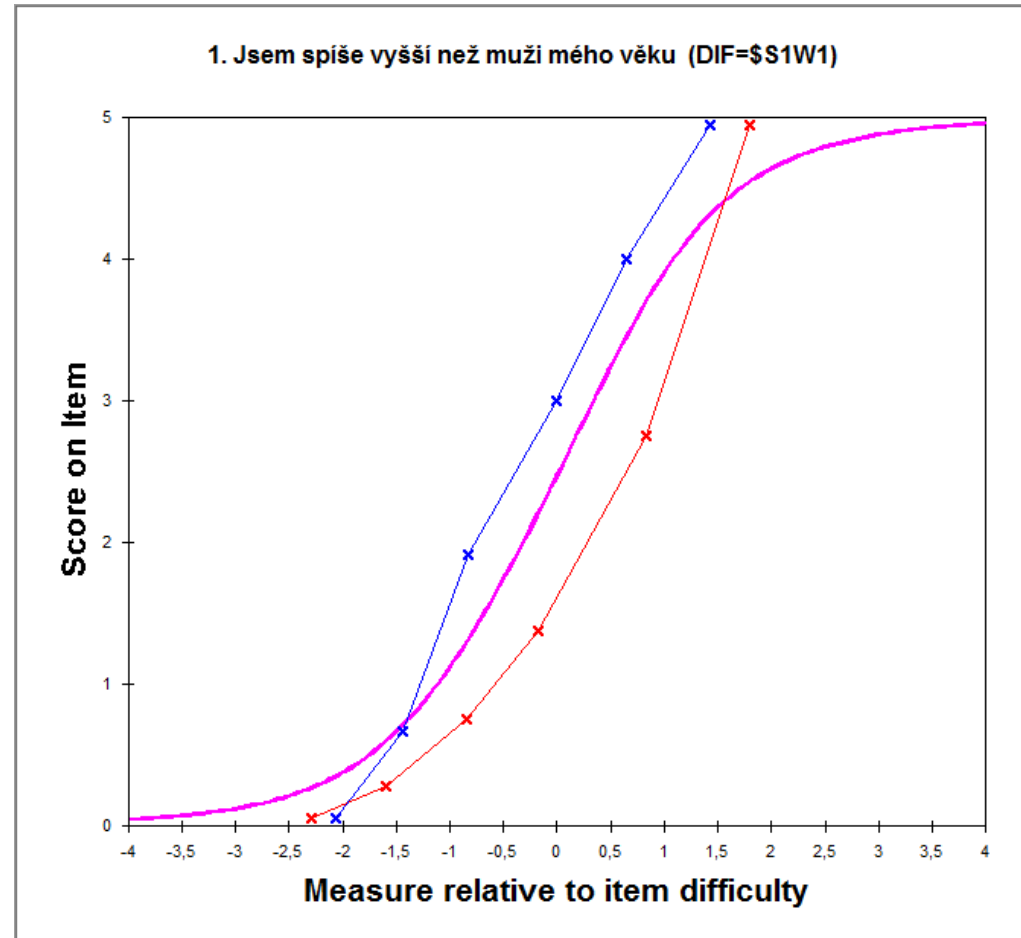
Dotazník výšky:

*Jsem spíše vyšší než muži mého věku.*

DIF:

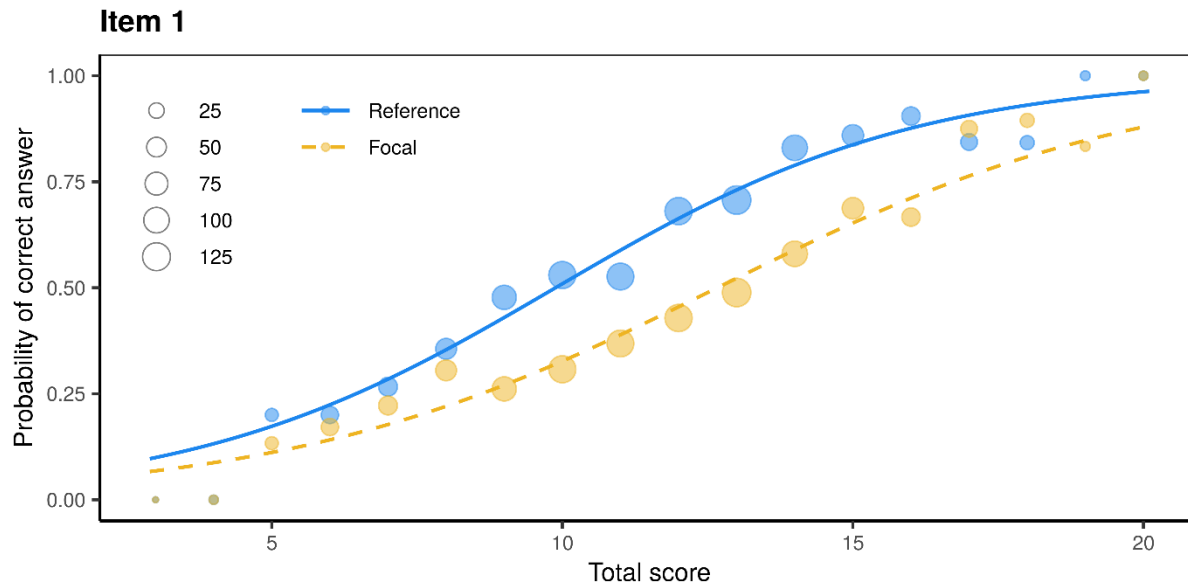
- t-test:  $t(86) = -4,63$ ,  $p < 0,001$
- M-H:  $\chi^2(1) = 18,7$ ,  $p < 0,001$ .

Modrá muži, červená ženy.

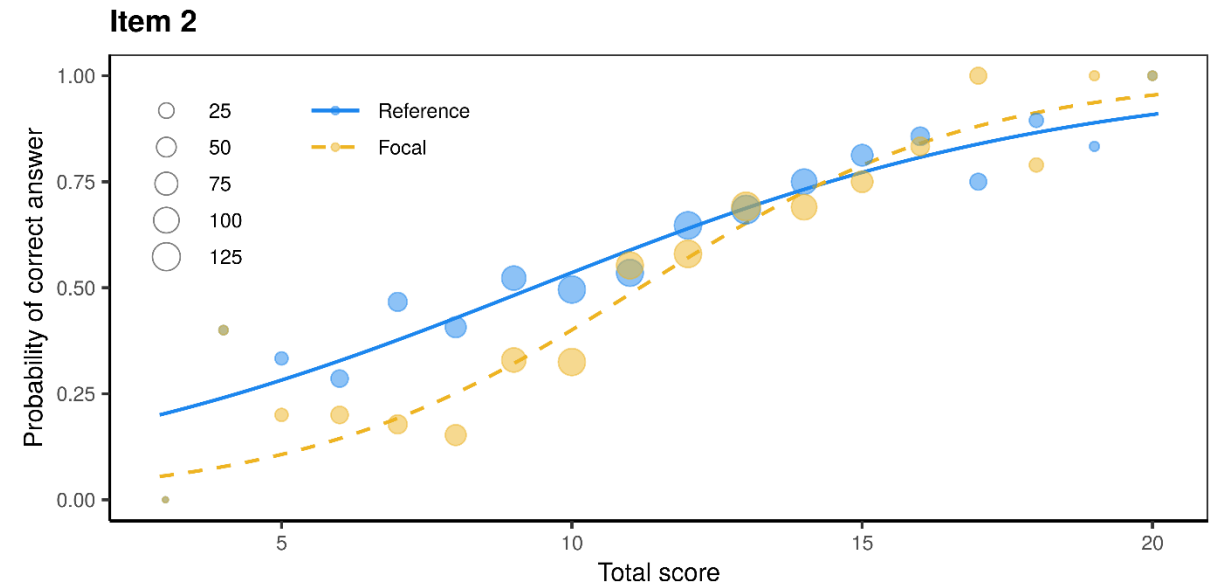


# Příklad: uniform vs. non-uniform DIF

## Uniformní DIF



## Non-uniformní DIF



# IRT přístupy k DIF

---

## Srovnání modelů (Model Comparison)

- Odhad multigroup IRT modelu (MG IRT) se stejnými koeficienty napříč skupinami.
- Odhad MG IRT modelů s uvolněnými koeficienty separátně pro každou položku.
- Modely jsou srovnány (LRT test, změna indexů přibližné shody modelu s daty – TLI, RMSEA, BIC...).
- Signifikance modelu znamená DIF, velikost efektu lze vyjádřit rozdílem shody modelu s daty nebo velikostí rozdílu parametrů.
- **Bottom-up:** východiskem je konfigurální model, fixování koeficientů (model se nesmí zhoršit).
- **Top-down:** východiskem je restriktivní model, uvolňování koeficientů (model se nesmí zlepšit).

## Explanační IRT modely, LLTM modely atp.

- Podobné přístupu založenému na logistické regresi.
- K položkám jsou doplněny vysvětlující skupinově-specifické proměnné.
- Pozornost je věnována signifikanci těchto proměnných na úrovni jedné položky.

Mnoho testů → LASSO, EBICglasso, korekce proti opakovanému testování atd.

# Software

---

Zejména R, různé balíčky

- difNLR package, mirt, difR, lordif, DIFlasso, DIFtree...

On-line aplikace: <https://shiny.cs.cas.cz/ShinyItemAnalysis/>

Jakýkoli statistický program, který disponuje modulem pro (ordinální) logistickou regresí.

# Invariance měření: Férovost z hlediska psychometrie.

---

KONGIGURÁLNÍ, METRICKÁ, SKALÁRNÍ

# Test bias: Invariance měření

---

Zkreslení na úrovni testu není nutné ověřovat jen pomocí prediktivní validity.

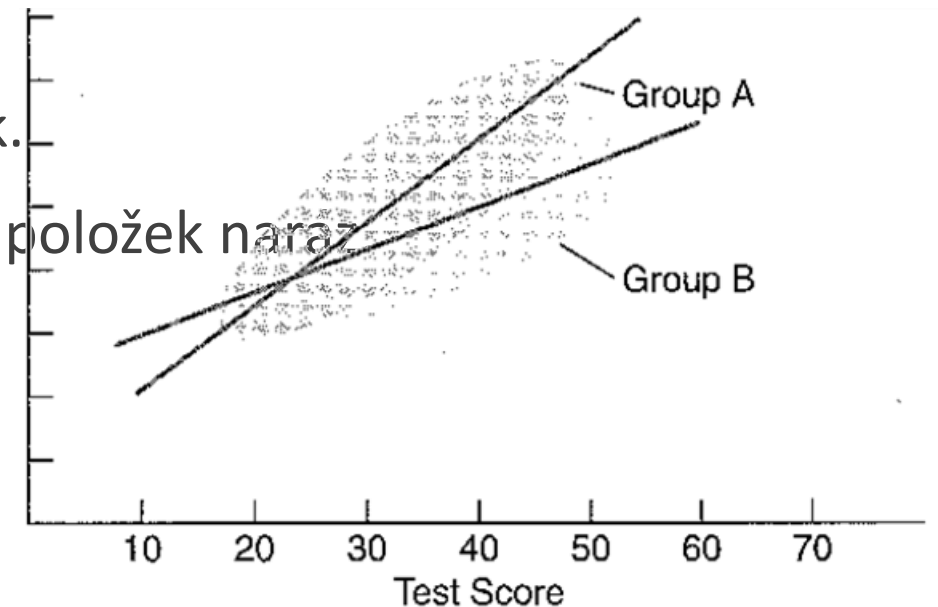
„Měří test stejný konstrukt napříč skupinami“?

Pokud ano, vztah odpovědí a úrovně latentního rysu by měl být shodný.

- A tedy i parametry položek by měly být shodné.

DIF se zaměřuje na parametry jednotlivých položek.

**Invariance měření** sleduje shodu všech parametrů položek naraz



# Test bias: Invariance měření

---

Postup založený na konfirmační faktorové analýze, ale je použitelný i v IRT.

- Tzv. multiple-group CFA/IRT (MG CFA, MG IRT).

Ověřuje shodnost faktorové struktury (modelu měření) napříč skupinami.

- Rozdílné úrovně invariance umožňují rozdílné možnosti srovnání skupin.

Typicky se řeší při:

- Při konstrukci diagnostických metod: je test jako celek „férový“ pro různé skupiny respondentů?
- Large-scale assessment: Do jaké míry mohu srovnávat skóry respondentů napříč zeměmi/státy/kulturami atd.?
- Teoreticky při každém použití t-testu by měla být vyargumentovaná invariance napříč oběma skupinami, aby je bylo možné srovnat.



# 4 (5) stupňů invariance:

---

## Základní:

- 1. Konfigurální invariance.
- 2. Metrická (slabá invariance).
- 3. Skalární (silná invariance).

## Další:

- 4. Reziduální (striktní) invariance.
- 5. Paralelní skupiny.

## Jednotlivé stupně/úrovně:

- Vyšší úrovně zahrnují všechny požadavky úrovní nižších.
- Nižší úrovně jsou předpokladem úrovní vyšších.

## Analogie k „paralelním položkám.“

- Paralelní položky: srovnání různých položek navzájem uvnitř jedné skupiny.
- Invariance: srovnání stejných položek napříč skupinami.

# 4 (5) stupňů invariance:

---

## 1. Konfigurální invariance:

- Test má stejnou strukturu (počet faktorů, přiřazení položek faktorům atd.) napříč skupinami.
- Měří tedy obsahově „ty stejné rysy“, ale klidně úplně „jinak“.
- Přesná definice rysů se může mírně lišit.
- Nelze srovnávat M a SD napříč skupinami, měřítko metody je jiné.

## 2. Metrická (slabá) invariance:

- Faktorové náboje v CFA jsou shodné (intercepty se mohou lišit).
- „Definice“ latentního rysu je stejná, má „stejné měřítko“, ale referenční bod je odlišný.
- Umožňuje srovnávat korelace latentních skóru napříč skupinami apod.
- Analogie tau-ekvivalentních položek.

# 4 (5) stupňů invariance:

---

## 3. Skalární (silná) invariance

- Intercepty v CFA jsou stejné napříč skupinami.
- Umožňuje srovnávat průměry latentních skóre skupin či respondenty napříč skupinami.
  - Např.: Češi mají vyšší skóre v PISA testech než Slováci.
  - Např.: Pacienti v dotazníku dosahují nižšího skóre než neklinická populace.
- Analogie paralelních položek.
- V tomto případě má prostý součet položek stále trochu jiný „význam“ (kvůli rozdílným reziduálním rozptylům).
  - Lze ale zanedbat, má vliv jen na signifikanci srovnání skupin a velikost efektu, nikoliv na „možnost“ takového srovnání.

# 4 (5) stupňů invariance:

---

## 4. Reziduální (striktní) invariance

- Položky mají v CFA modelu stejný chybový rozptyl.
- Analogie striktně-paralelních položek.
  - Vztah součtu položek a latentního rysu je napříč skupinami stejný.

## 5. Paralelní skupiny

- Na rozdíl od předchozího není vlastností jen testu, ale i skupiny.
- Jednotlivé skupiny respondentů mají stejné průměry a rozptyly.
- Jinými slovy, neexistuje žádný pozorovatelný rozdíl napříč skupinami v odpovědích na test.

# Typické stupně invariance

- Alternativně lze fixovat vybraný faktorový náboj, nikoliv lat. rozptyl.
- Pořadí není zcela pevně dané, jen 1. a 2. krok jsou nezbytné pro všechny další;
- Krok 3 je předpokladem pro 5a a 6; 5a a 5b lze přeskočit a rovnou testovat 6.
- Pozor, v ordinální CFA a v IRT jsou určité odlišnosti!

	náboje	intercepty	rezidua	lat. průměry	lat. rozptyly
<b>1. konfigurální</b>	volné	volné	volné	fixované (0)	fixované (1)
<b>2. metrická (slabá)</b>	omezené	volné	volné	fixované (0)	ref. skup. fixované (1) další skup.: volné
<b>3. skalární (silná)</b>	omezené	omezené	volné	ref. skup. fixované (0) další skup.: volné	ref. skup. fixované (1) další skup.: volné
<b>4. reziduální (striktní)</b>	omezené	omezené	omezené	ref. skup. fixované (0) další skup.: volné	ref. skup. fixované (1) další skup.: volné
<b>5a. ekvivalence průměrů</b>	omezené	omezené	omezené	fixované (0)	ref. skup. fixované (1) další skup.: volné
<b>5b. ekvivalence rozptylů</b>	omezené	omezené	omezené	ref. skup. fixované (0) další skup.: volné	fixované (1)
<b>6. ekvivalentní skupiny</b>	omezené	omezené	omezené	fixované (0)	fixované (1)

# Alternativní způsoby ověření invariance

---

Multi-group CFA není jediným postupem.

Přehled všech postupů předkládá [Kim, Cao, Wang and Nguyen \(2017\)](#).

- Multiple group confirmatory factor analysis (MG CFA).
- Multilevel confirmatory factor analysis (ML CFA).
- Multilevel factor mixture modeling (ML FMM).
- Bayesian approximate M.I. testing (using BSEM).
- Alignment optimization.

Není nutné znát. MG CFA je zlatý standard a v psychologii postačuje.

- Potíž nastává při velkém množství skupin, kdy jsou alignment, ML CFA a BSEM výhodnější (typicky v mezinárodních ILSA studiích).

# Invariance: Další témata

---

## Invariance v IRT a ordinální CFA

- Typicky je komplikovanější odlišení metrické a skalární invariance.
- Řada různých parametrizací, řada zádrhelů.

## Modifikační indexy invariantní MG CFA modelu

- Analogie top-down DIF analýzy.

## Longitudinální invariance

- Měří test ten samý rys u těch stejných respondentů v průběhu času?
- Vývojová psychologie, intervence, terapie...

## Využití invariance a DIF při vyvažování paralelních forem

- Invariance kotevních testů, DIF kotevních položek.

## Invariance jako nedílná součást ILSA (International Large Scale Assessment)

- Dost krize.

# Doporučený postup

(Cígler, personal communication 😊)

---

Během vývoje testu: průběžné DIF analýzy (souběžně s položkovými analýzami) pro ověření kvality položek.

Během standardizace:

- Ověření invariance metody.
- Pokud je non-invariantní, pak DIF analýza pro identifikaci problémových položek.
- Vyřazení problémových položek, případně úprava skórování.
- U raschovských metod někdy agregace DIF analýz namísto analýzy invariance (WJ IV us).

Během vyvažování paralelních forem testu či adaptace.

- Kombinace analýz DIF a invariance.

Po standardizaci za účelem validizace.

- Dodatečné analýzy invariance pro skupiny nezahrnuté do standardizační studie.



# Take-home message

---

1. Férovost (fairness) je zavedený termín v oblasti validity.
2. Diagnostická vs. psychometrická rovina férovosti.
3. Nelze srovnávat lidi napříč různými skupinami bez dostatečné empirické podpory.
4. Dobrá diagnostická metoda by měla poskytovat výsledky analýz invariance a DIF napříč smysluplnými populacemi.
5. Zvážení férovosti je nezbytnou součástí psychologické diagnostiky.  
„Dodal mi vydavatel testu dostatek informací“?  
„Cítím dostatečnou podporu pro validitu ve smyslu důsledků testování“?
6. Termíny: DIF, invariance, response/item/test/predictive bias, accesibility, universal design, akomodace/modifikace.



**A DAY MAY COME WHEN  
DIF ANALYSIS IS NOT NEEDED**

**BUT NOT TODAY**