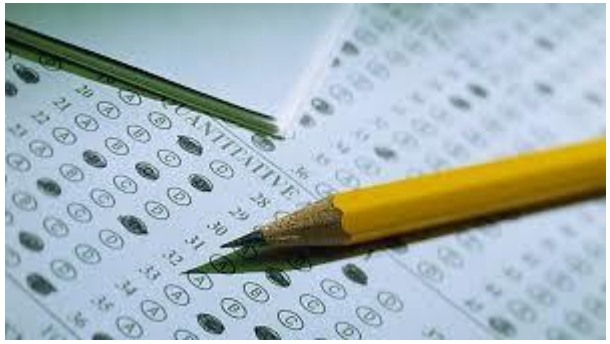
VALIDATION OF CONCEPTUAL ASSESSMENTS AND ADMISSION TESTS FOR MEDICAL SCHOOLS



Patrícia Martinková
Dept. of Medical Informatics and Biostatistics
Institute of Computer Science
Czech Academy of Sciences
martinkova@cs.cas.cz
www.cs.cas.cz/martinkova

Mefanet 2016, November 30, Brno

- (2015-2017) Grant GAČR GJ15-15856Y: Odhad psychometrických vlastností jako součást vývoje přijímacích testů

- (2013-2015) pobyt na University of Washington, Seattle
  - Fulbright-Masarykovo stipendium
  - Přednášky Item-Response Theory Models
  - Přednášky základů statistiky
  - Účast na IES projektu analýzy přijímacích řízení učitelů
  - Účast na NSF projektu analýzy tzv. conceptual assessments

Obsah prezentace

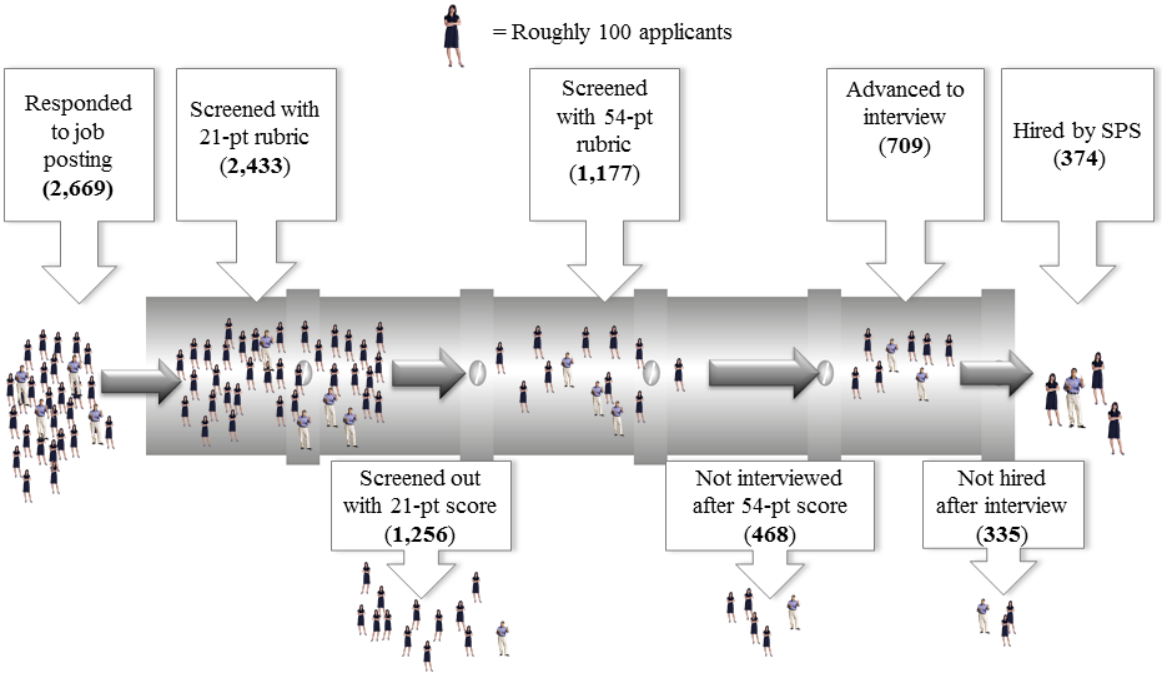
Kombinace dvou témat:

- Validizace **přijímacích zkoušek** na lékařské fakulty
- Validizace tzv. **conceptual assessments** (concept inventories)

V souhrnu: **jak testovat znalosti co nejlépe**

Přijímací řízení

- **Klíčový moment** pro studenta
- **Klíčový moment** pro instituci (lékařskou fakultu)



Hiring process in Spokane public schools, Martinkova & Goldhaber, 2015

Přijímací řízení na lékařských fakultách

Inspirace ze zahraničí:

- **MCAT** - Medical College Admission Test
  (připravovaný American Association of Medical Colleges)
- **UMAT** – Undergraduate Medical College Admission Test
- **GAMSAT** – Graduate Australian Medical School Admission Test
- …

Přijímací řízení na lékařských fakultách

Inspirace ze zahraničí:

- Rutinní komplexní systém přípravy přijímacích zkoušek
  - Design testu (blueprint): témata, typy úloh
  - Příprava a následná recenze položek
  - Zabezpečení položek
  - Analýza testu a položek … KOMPLEXNÍ VALIDIZACE TESTU

- Výhody pro univerzitu i studenta
  - Ještě lepší testovaní a výběr studentů
  - Možnost více termínů, administrace

Validizační studie přijímacích zkoušek na LF

- Štuka, Martinková et al. (2012)

  Predikují výsledky v testu na 1. LF UK budoucí úspěšnost studenta?

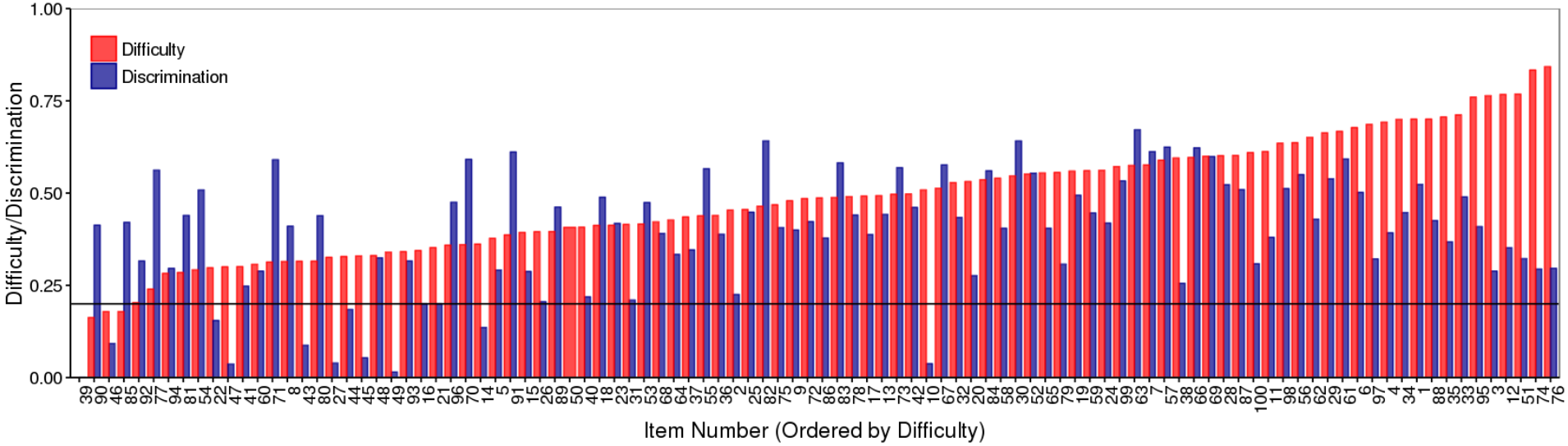
  - Úspěšnost: dostudování LF do 7 let

  - Model, v němž úspěšnost závisí pouze na **SŠ známkách** vysvětlí úspěšnost z 15%
  - Model, v němž úspěšnost závisí pouze na **počtu bodů u přijímaček** vysvětlí úspěšnost z 13%
  - Model, v němž úspěšnost závisí na **SŠ známkách i na počtu bodů** vysvětlí úspěšnost z 22%
  - Přidáme-li do modelu typ školy a rok maturity… model vysvětlí úspěšnost z 31%

  Závěr: Počet bodů u přijímaček je **signifikantní** prediktor úspěšnosti, přidává navíc novou informaci nad SŠ prospěch.

Validizace přijímacích zkoušek na LF

- Podpora tvorby kvalitních položek
- Analýza predikční schopnosti testu a jeho položek, optimalizace nastavení parametrů přijímací zkoušky…
- Analýza položek



Podpora analýzy přijímacích zkoušek na LF

https://shiny.cs.cas.cz/ShinyItemAnalysis/

TEST AND ITEM ANALYSIS

Item

1 20

1 3 5 7 9 11 13 15 17 19 20

ShinyItemAnalysis Version 0.2

Version 0.1 is available here.

Project was supported by grant funded by Czech Science foundation under number GJ15-15856Y

Copyright 2016 Patricia Martinkova, Ondrej Leder and Adela Drabinova

Hits:95

Total Scores | Item Analysis | Distractors | Regression | IRT | DIF/Fairness | Data | About

ShinyItemAnalysis provides analysis of educational tests (such as admission tests) and its items. For demonstration purposes, 20-item dataset **GMAT** from R `library(difNLR)` is used. You can change the dataset (and try your own one) on page **Data.**

Analysis of Total Scores

Summary Table

Min	Max	Mean	Median	SD	Skewness	Kurtosis
3	19	11.58	12	3.08	-0.1	2.55

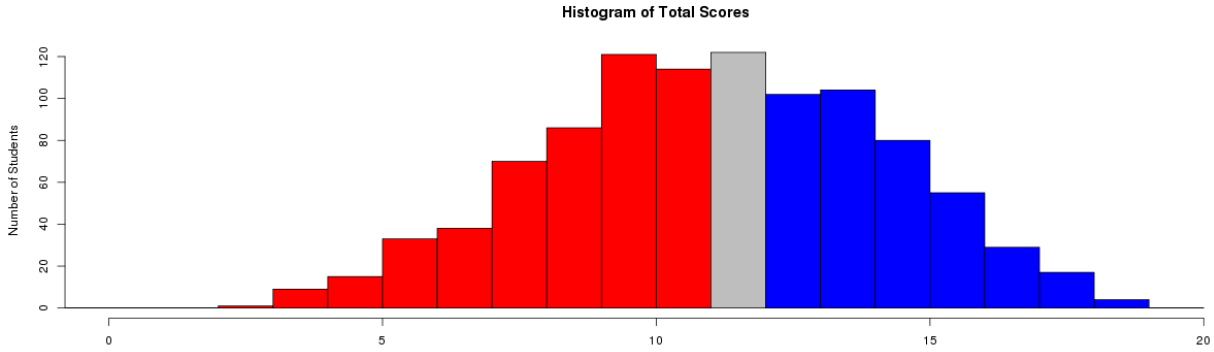
Histogram of Total Score

Cutpoint

0 12 20

0 2 4 6 8 10 12 14 16 18 20

For selected Cutpoint, blue part of histogram shows students with total score above the cutpoint, grey column shows students with Total Score equal to cutpoint and red part of histogram shows students below the cutpoint.



Validizace v obecnějším pojetí

Důkazy validity - empirické:

- **Predikční:** Počet bodů u přijímaček predikuje budoucí úspěšnost studenta
- **Inkrementální:** Počet bodů u přijímaček vylepšuje predikci budoucí úspěšnosti studenta nad tu obsaženou v jiných dostupných údajích (maturita, prospěch na SŠ)
- **Diskriminační:** Test dobře rozlišuje mezi skupinami, mezi kterými by měl rozlišovat
- **Souběžná:** Výsledky v testu korelují s výsledky v jiných testech

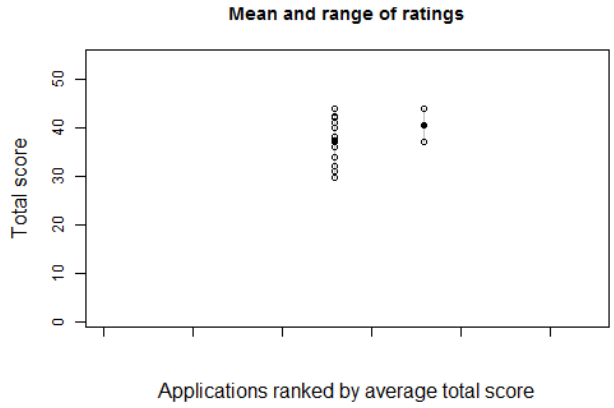
Důkazy validity - obsahové:

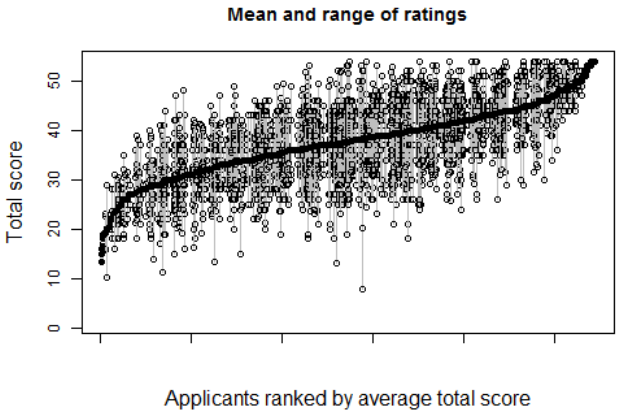
- Definice testované oblasti, rozdělení domény do kategorií, posouzení počtu položek v kategoriích (**blueprinting**)

Validizace v obecnějším pojetí

Reliabilita - test „není zatížený chybami“

- **Test-retest reliabilita:** Výsledky se nezmění, pokud měření provedeme (nezávisle) znovu
- **Vnitrotřídní konzistence** (Cronbachovo alfa): Položky měří stejný konstrukt
- **Inter-rater reliabilita:** hodnocení jednotlivých hodnotitelů (nebo komisí) jsou konzistentní





Validizace v obecnějším pojetí

Analýza položek

- Obtížnost
- Diskriminační schopnost
- Analýza distraktorů
- Férovost položek

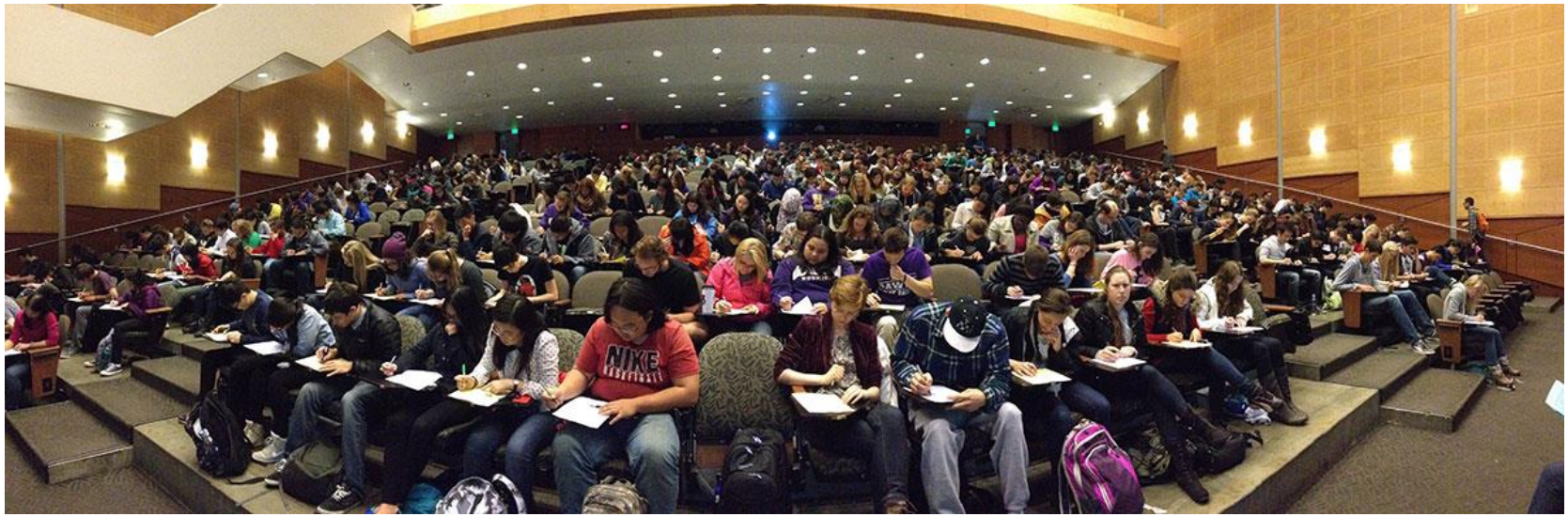
Conceptual Assessments in Physiology

http://physiologyconcepts.org/

Conceptual Assessments in Physiology

http://physiologyconcepts.org/

- NSF projekt, McFarland et al.



http://uwberg.com/

Biology Education Research Group, University of Washington

Conceptual Assessments in Physiology

http://physiologyconcepts.org/

- NSF projekt, McFarland et al.
- Motivace:
  - Velký objem učiva
- Cílem:
  - Zlepšit výuku (undergraduate) fyziologie strukturováním učiva
  - Definovat nejdůležitější koncepty fyziologie (identifikace15 hlavních konceptů)
  - Dále strukturovat identifikované koncepty
  - Vytvořit testy hodnotící znalost hlavních konceptů: **Conceptual Assessments (Concept Inventories)**

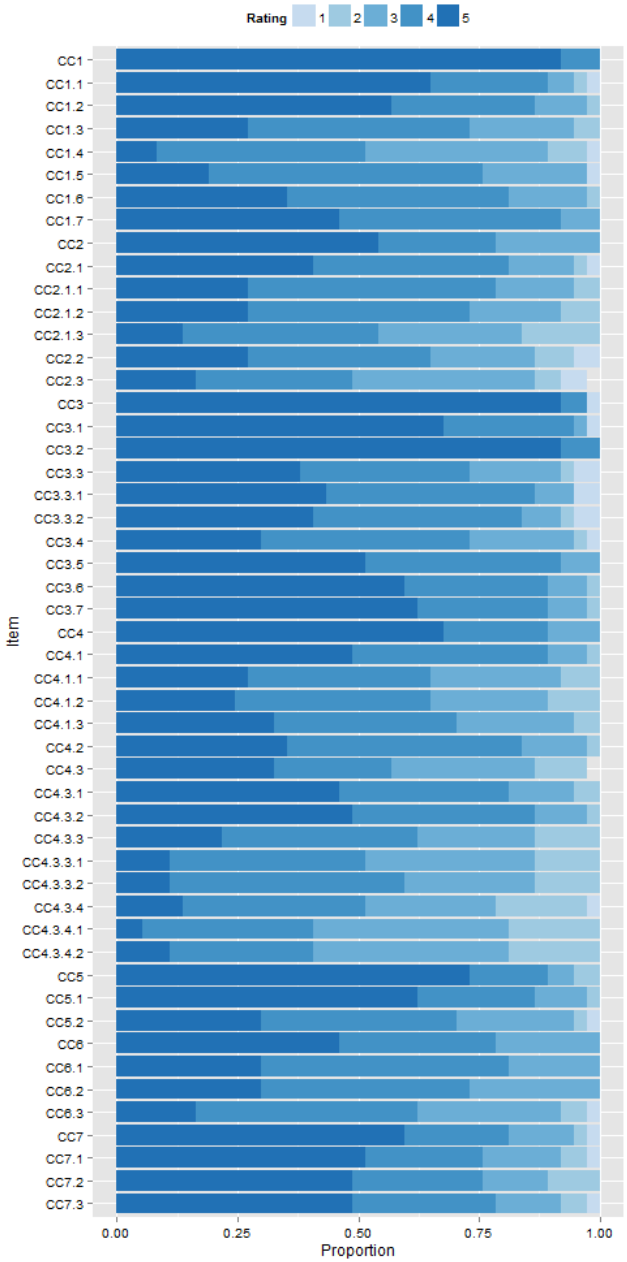
Conceptual Assessments

Příklad 1:

Cell-cell Communication (CCC) Framework

(Michael, Martinková, McFarland et al, *Advances in Physiology*, in review)

- **CC1.** Chemical Messenger Synthesis & Release
- **CC2.** Chemical Messenger Transport
- **CC3.** Receptor Binding
- **CC4.** Signal Transduction
- **CC5.** Cell Response
- **CC6.** Signal Termination
- **CC7.** Electrical Communication



Conceptual Assessments

Příklad 2:

Validation of Homeostasis Concept Inventory (HCI)

(McFarland, Price, Wenderoth, Martinkova et al, *CBE LSE*, in review)

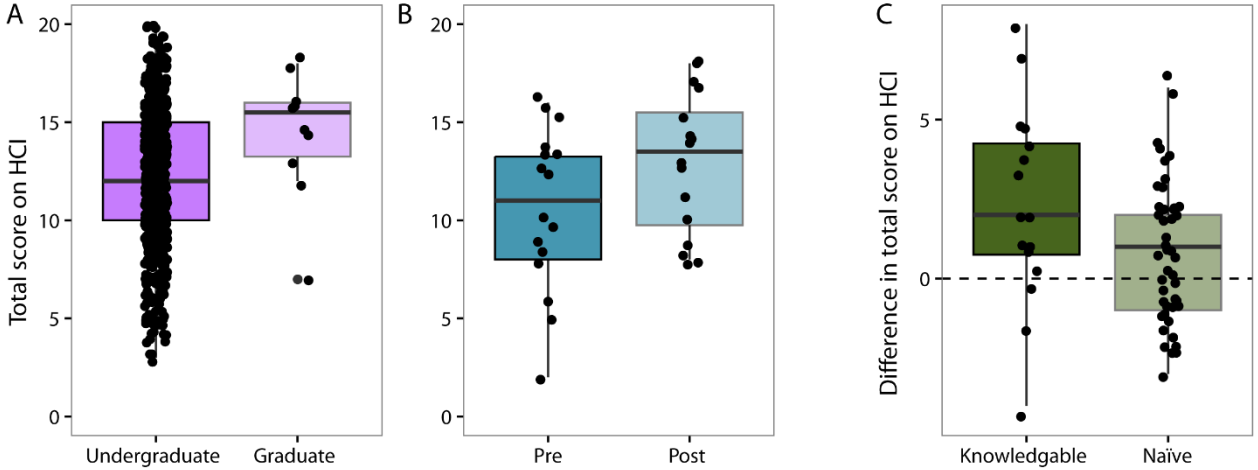
- Navazuje na Homeostasis Conceptual Framework (HCF)
- 20-položkový znalostní test
- Komplexní validizace testu:
  - Obsahová validita
  - Predikční validita
  - Reliabilita
  - Analýza položek

Conceptual Assessments

Příklad 2:

Validation of Homeostasis Concept Inventory (HCI)

(McFarland, Price, Wenderoth, Martinkova et al, *CBE LSE*, in review)

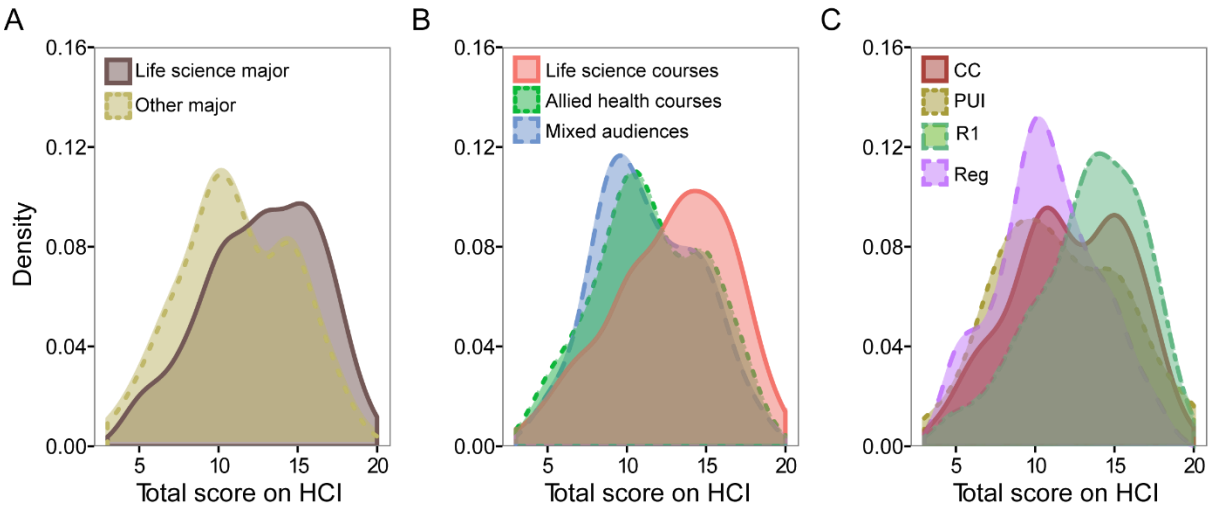


Conceptual Assessments

Příklad 2:

Validation of Homeostasis Concept Inventory (HCI)

(McFarland, Price, Wenderoth, Martinkova et al, *CBE LSE*, in review)

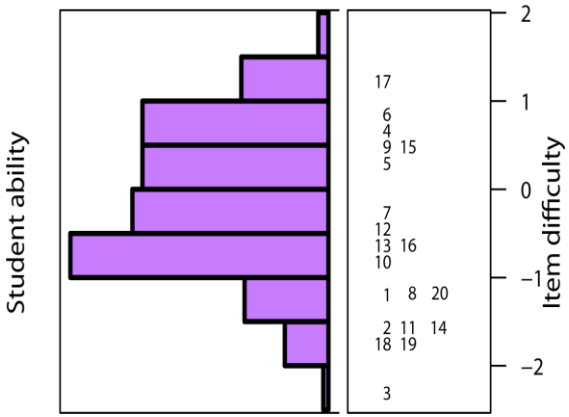


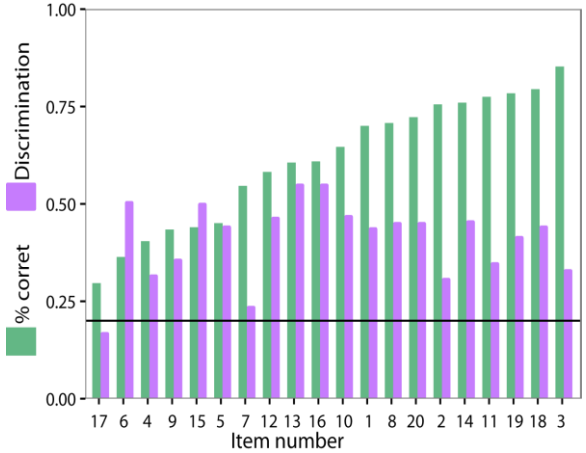
Conceptual Assessments

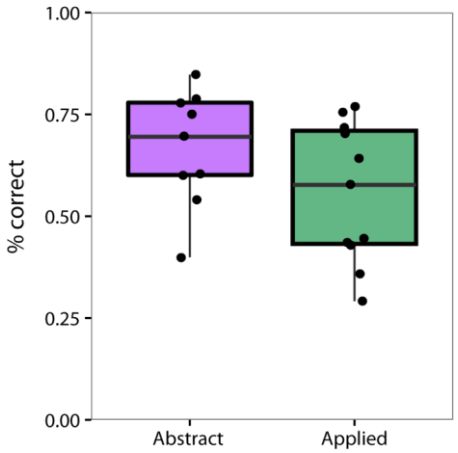
Příklad 2:

Validation of Homeostasis Concept Inventory (HCI)

(McFarland, Price, Wenderoth, Martinkova et al, *CBE LSE*, in review)





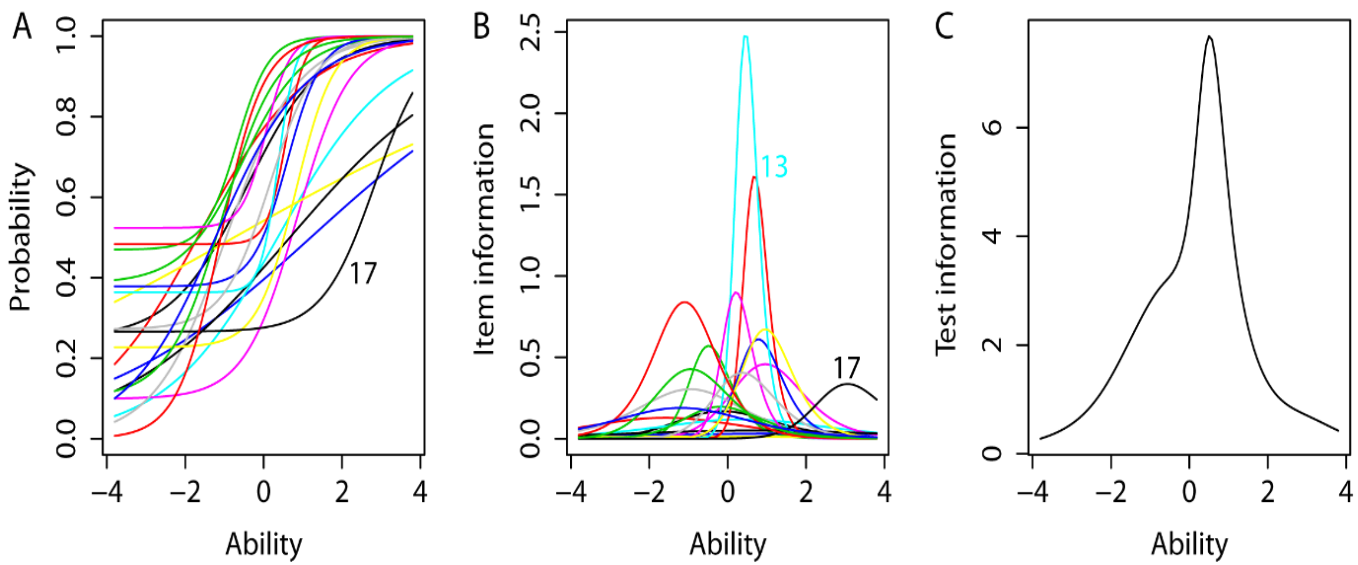


Conceptual Assessments

Příklad 2:

Validation of Homeostasis Concept Inventory (HCI)

(McFarland, Price, Wenderoth, Martinkova et al, *CBE LSE*, in review)



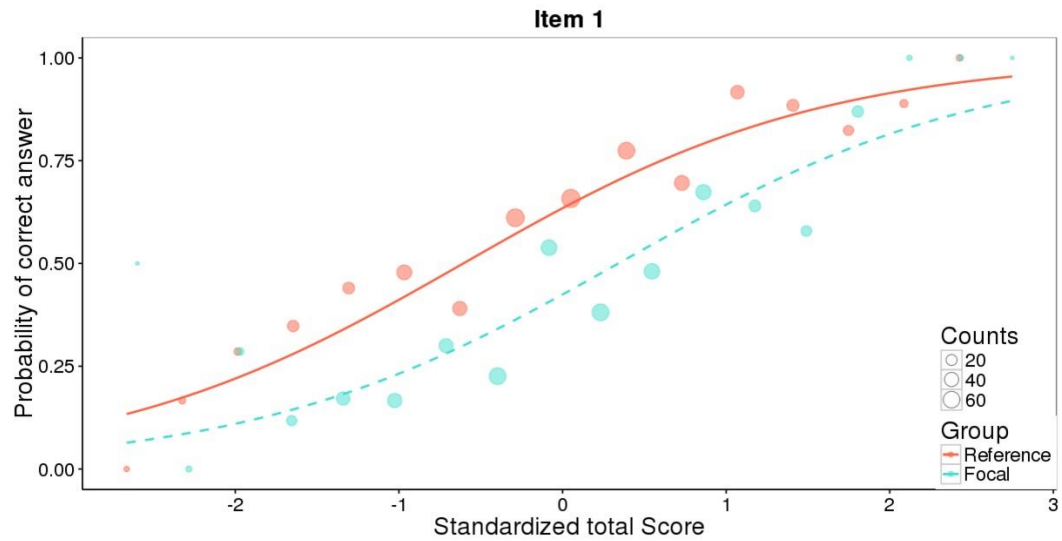
Conceptual Assessments

Příklad 3. Odlišné fungování položky pro skupiny

DIF analysis to reveal equity gaps in conceptual assessments
(Martinková, et al., *CBE LSE*, in review)

Př.: Nedostatek vitamínu D v dětství vede ke:

a) **Křivici**
b) Kurdějím
c) Trpasličímu vzrůstu
d) Debilitě



Equation

$$\mathrm{P}\left(Y_{ij}=1|Z_i,G_i,a_j,b_j,a_{\mathrm{DIF}j},b_{\mathrm{DIF}j}\right)=\frac{e^{(a_j+a_{\mathrm{DIF}j}G_i)(Z_i-(b_j+b_{\mathrm{DIF}j}G_i))}}{1+e^{(a_j+a_{\mathrm{DIF}j}G_i)(Z_i-(b_j+b_{\mathrm{DIF}j}G_i))}}$$

- Detekován DIF
- Je položka neférová vůči mužům?
  Tato nikoliv (dětské nemoci jsou relevantní znalostí pro LF)

Shrnutí pro Conceptual Assessments

Komplexní validizace znalostního testu:

- Obsahová validita: definování Framework (např. CCC, HCF),
- Validita predikční:
  - Test koreluje s celkovým prospěchem studenta
  - Test dobře rozlišuje mezi skupinami
- Reliabilita
  - test-retest reliabilita
- Detailní analýza položek
  - Obtížnost
  - Diskriminační schopnost
  - Dobře fungující distraktory
  - Identifikace tzv. misconceptions

Závěr a diskuse

- Kvalitní Conceptual Assessments jsou tu pro Vás
- Z dostupných validizačních studií se lze inspirovat i pro další testování (tedy i přijímací testy)

- Testování je klíčové a je nutné dělat jej kvalitně:
  - Přijímací řízení: okamžité důsledky pro studenta i instituci
  - Formativní hodnocení: špatné testy mohou odradit studenty od oboru

- **Spojení sil odborníků ==>**

  Kvalitní testy ==>

  Šťastní studenti i pedagogové

Děkuji za pozornost!

martinkova@cs.cas.cz
www.cs.cas.cz/martinkova/

https://shiny.cs.cas.cz/ShinyItemAnalysis/