mefanet 2012

MEDICAL FACULTIES NETWORK

E-testování:
Moderní trendy v hodnocení znalostí

RNDr. Čestmír Štuka, Ph.D., MBA (1. LF UK)

MUDr. Martin Vejražka, Ph.D. (1. LF UK)

Mgr. Martin Komenda (LF MU)

RNDr. Patrícia Martinková, Ph.D.(ÚI AV ČR)

MUDr. Jan Trnka, Ph.D. (3. LF UK)

Osnova workshopu

Životní cyklus testové agendy (10 min – Č. Štuka)

- Motivace pro realizaci workshopu + příprava společné publikace
- Představení základních fází přípravy testů (zjednodušené schéma)

Vytváření otázek (30 min – M. Vejražka)

- Ukázka jednotlivých typů otázek
- Tvorba SBA (prakticky ve skupinách)
- Oponentura (prakticky ve skupinách)

Standardizace (10 min – M. Komenda)

- Motivace, výhody a nevýhody
- Relativní a absolutní standardizace
- Podrobná ukázka standardizace: Angoffova metoda

Analýza výsledků testu (20 min – P. Martinková)

- Popis výsledků testu
- Reliabilita a validita testu
- Analýza položek

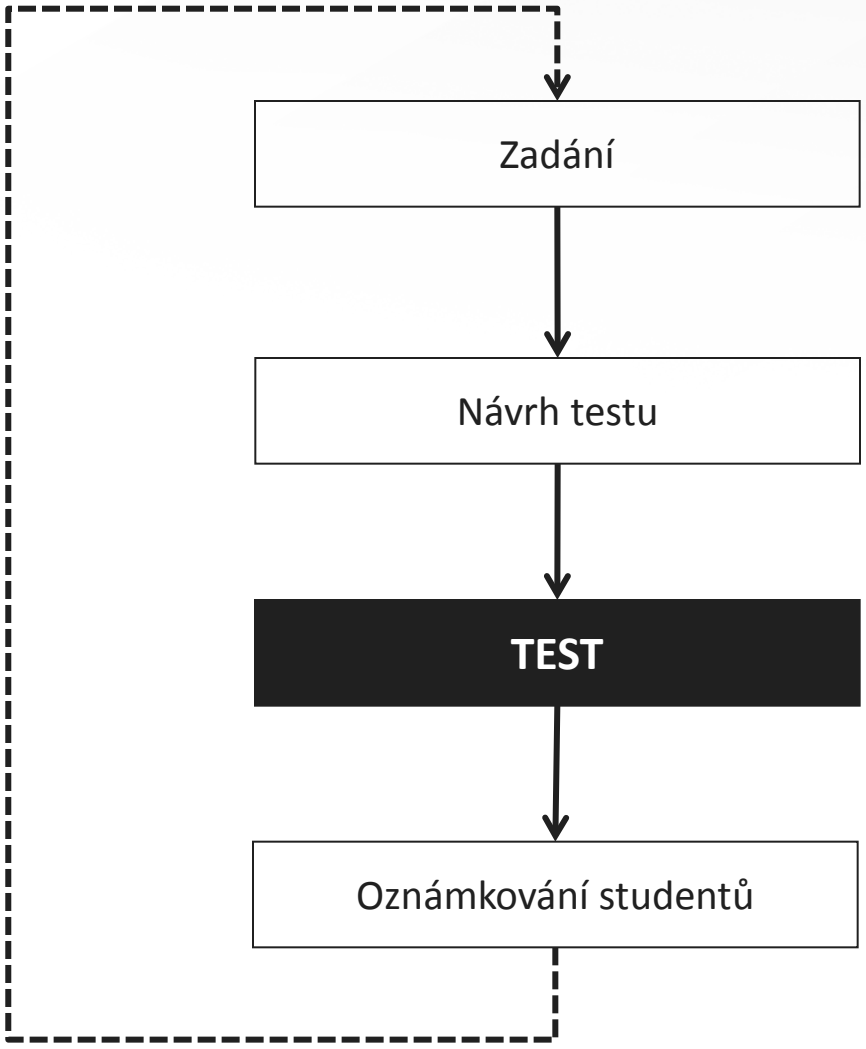
Závěrečná diskuse (10 min)

mefanet 2012

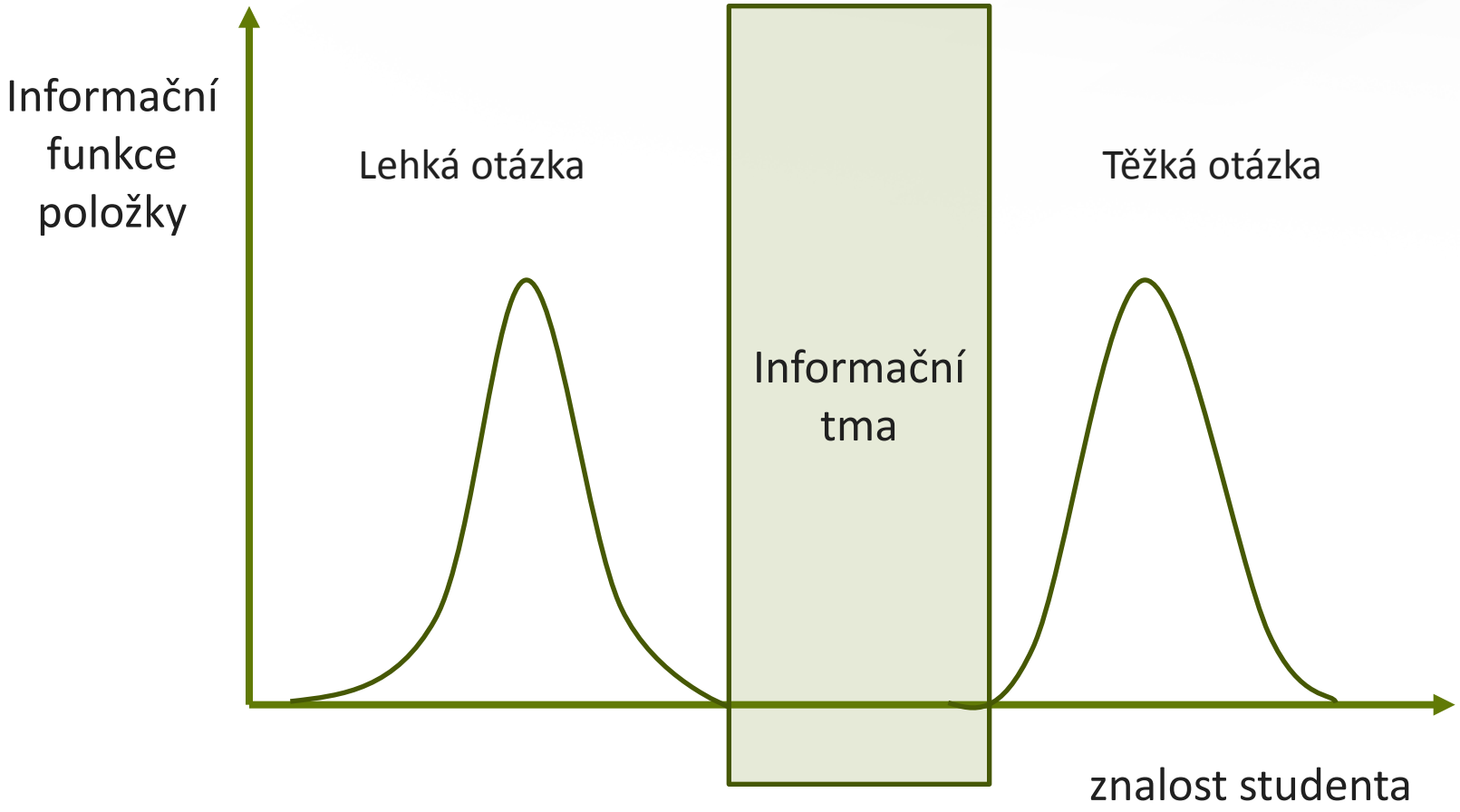
MEDICAL FACULTIES NETWORK

Životní cyklus testové agendy

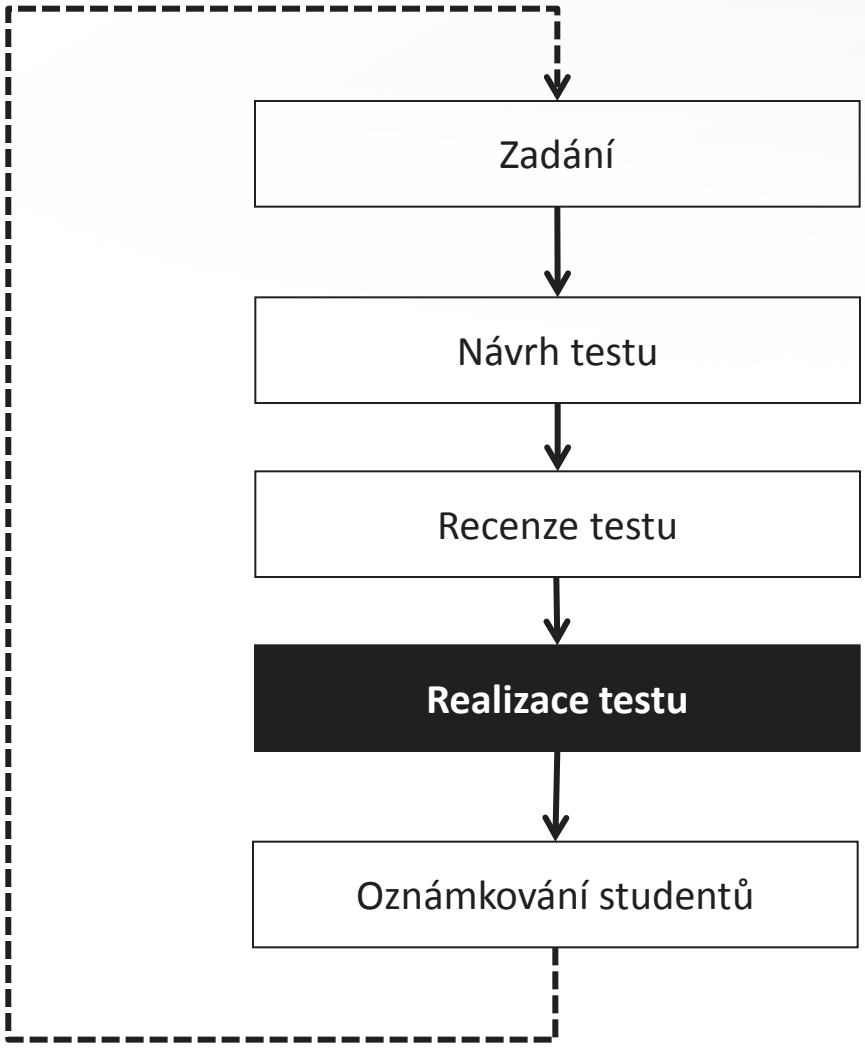
Nativní testový cyklus



Představme si test se dvěma otázkami



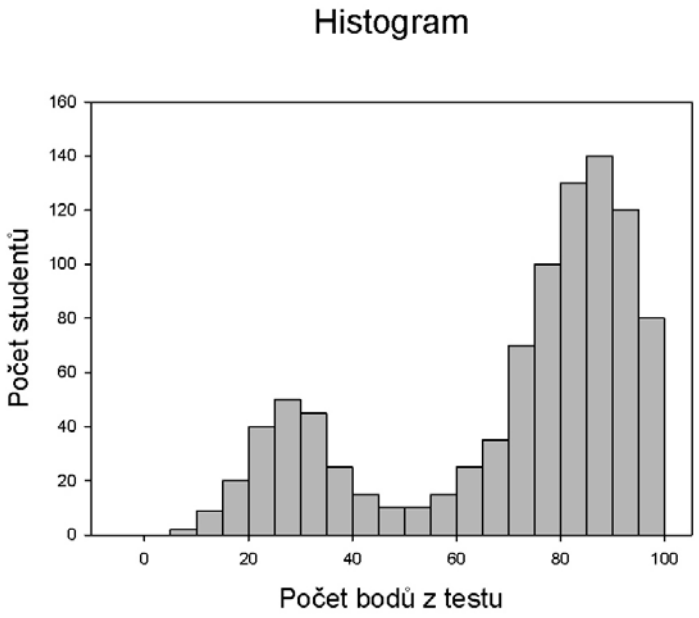
Zařazení předběžné recenze do testového cyklu



Zajímají nás výsledky testu ?

Máme důvod zabývat se výsledkem testu?

Poznáme „vynesený test“?

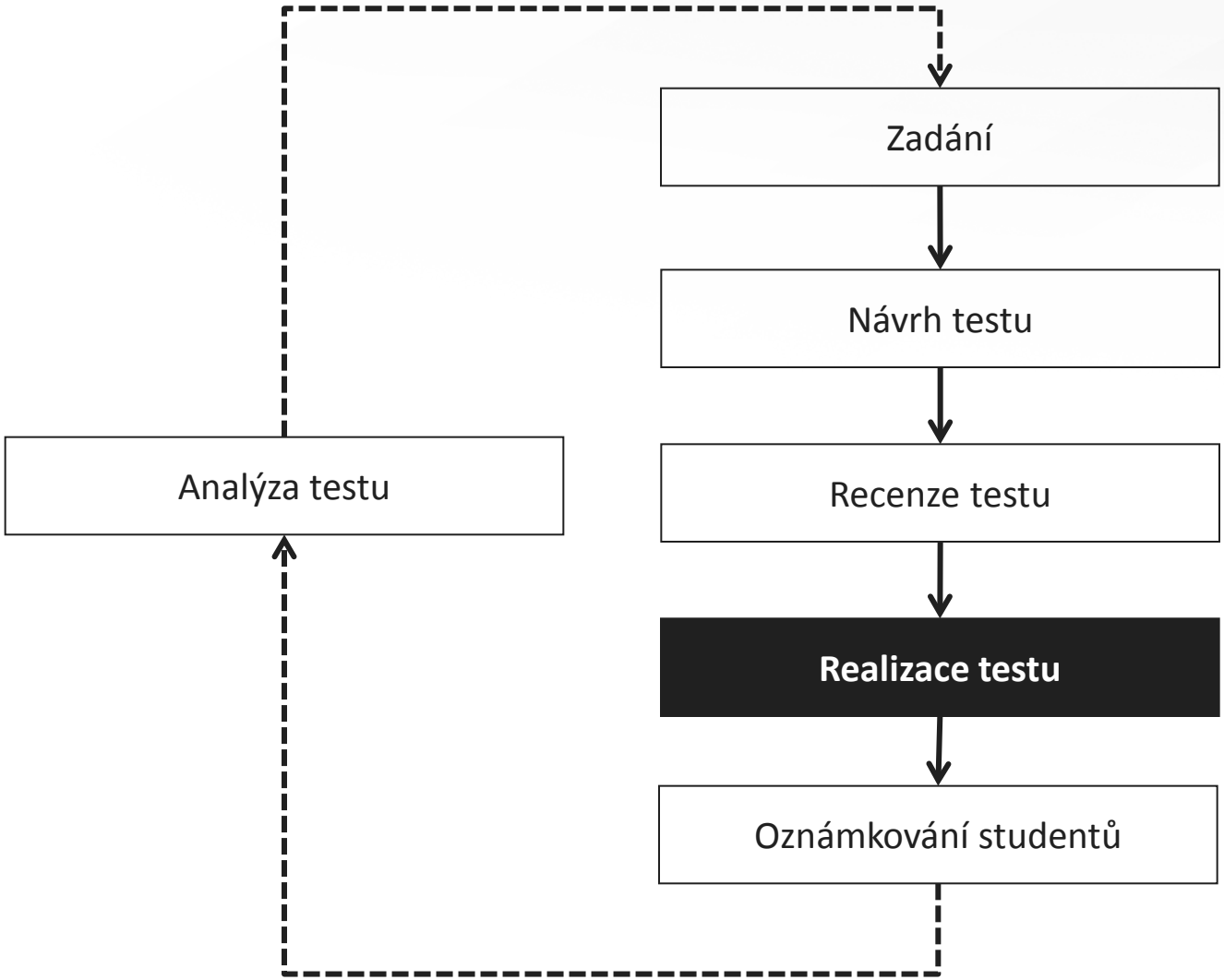


Poznáme „vynesenou položku“?

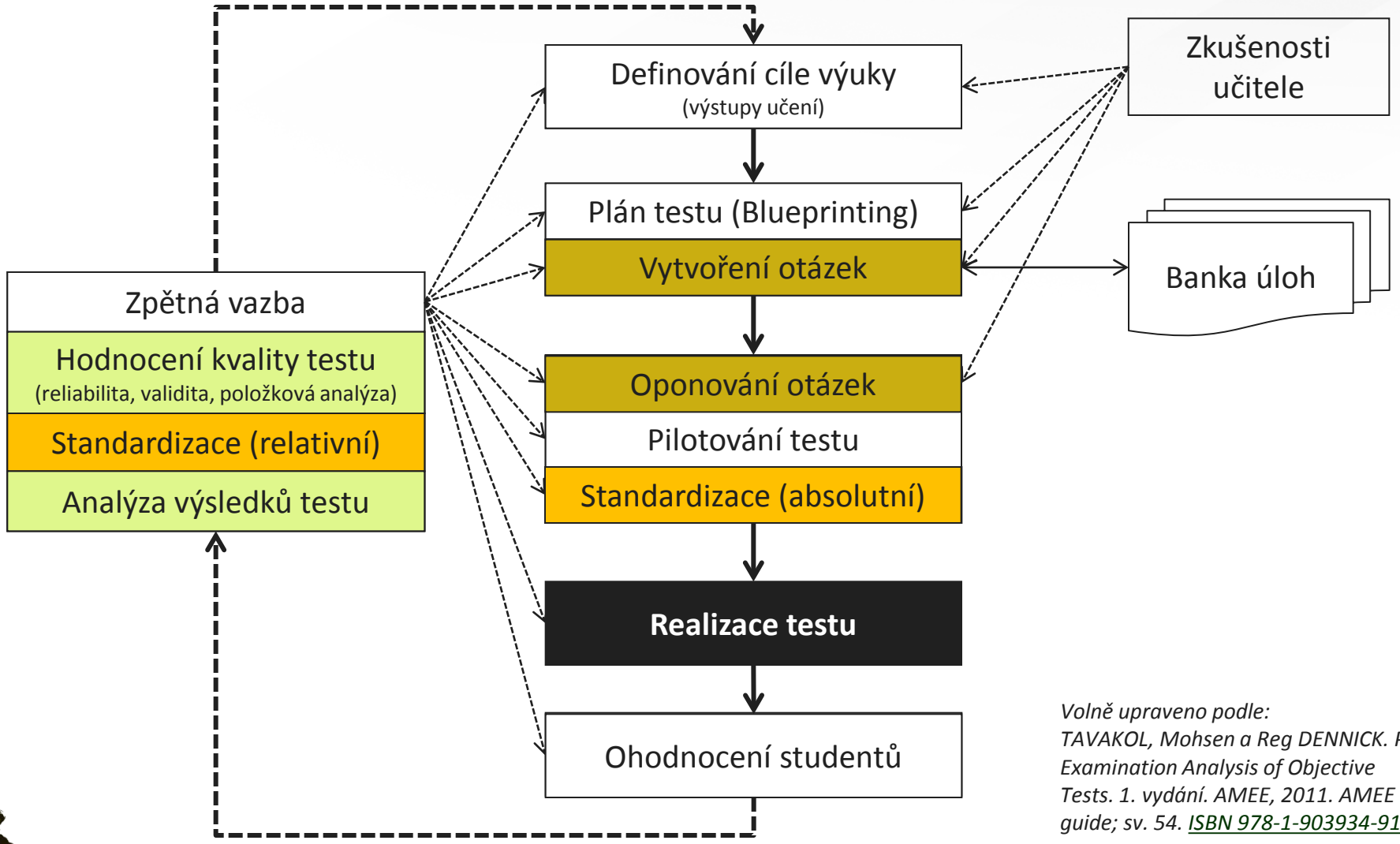
Změnila se skokově obtížnost položky mezi dvěma testy?

Zajímá nás ANALÝZA TESTU !

Vyhodnocení výsledků testu



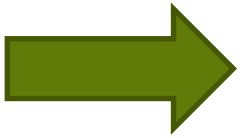
Testový cyklus



Volně upraveno podle: TAVAKOL, Mohsen a Reg DENNICK. Post Examination Analysis of Objective Tests. 1. vydání. AMEE, 2011. AMEE guide; sv. 54. ISBN 978-1-903934-91-3.

Kam dál?

Co nenajdeš zde, najdeš v lexikonu!



www.wikiskripta.eu/Testy

mefanet 2012

MEDICAL FACULTIES NETWORK

Vytváření otázek

Millerova pyramida

Písemné testování
NENÍ vhodné
- zvolte některou z forem praktického zkoušení

Písemné testování
JE vhodné

Otázky MTF

- Multiple true/false

Chlorid amonný

- ❑ má ve své molekule čtyři atomy vodíku
- ❑ rozpuštěním ve vodě dává zásaditý roztok
- ❑ uvolňuje po přidání hydroxidu sodného amoniak
- ❑ je silné oxidační činidlo

Chlorid amonný

má ve své molekule čtyři atomy vodíku	ANO / NE
rozpuštěním ve vodě dává zásaditý roztok	ANO / NE
uvolňuje po přidání hydroxidu sodného amoniak	ANO / NE
Je silné oxidační činidlo	ANO / NE

SAQ

- Otázky s krátkou tvořenou odpovědí (short-answer questions, SAQ)

Racionálními vzorci zapište azokopulační reakci benzendiazoniové soli s α-naftolem

- Nejsou vhodné pro automatizované testování
- Musí opravovat (vyškolený) odborník

Otázky SBA

- Otázky s jedinou nejlepší odpovědí (single best answer, SBA)

32letý muž přichází pro 4 dny trvající, postupně progredující slabost končetin. Dosud byl zdráv, před 10 dny však prodělal infekci horních cest dýchacích. Je afebrilní, arteriální tlak má 130/80 mmHg, tepovou frekvenci 94 / min. Dýchání je mělké a nápadně zrychlené. V orientačním neurologickém nálezu dominuje symetrická slabost mimických svalů a svalů horních i dolních končetin. Čití je intaktní. Hluboké šlachové reflexy nelze vybavit. Zánikové jevy jsou negativní. Která z následujících diagnóz je nejpravděpodobnější? (*Vyberte jedinou odpověď*)

- ❑ akutní diseminovaná encefalomyelitida
- ❑ syndrom Guillain-Barré
- ❑ myasthenia gravis
- ❑ poliomyelitis
- ❑ polymyositis

Otázky SBA

Pět zásad

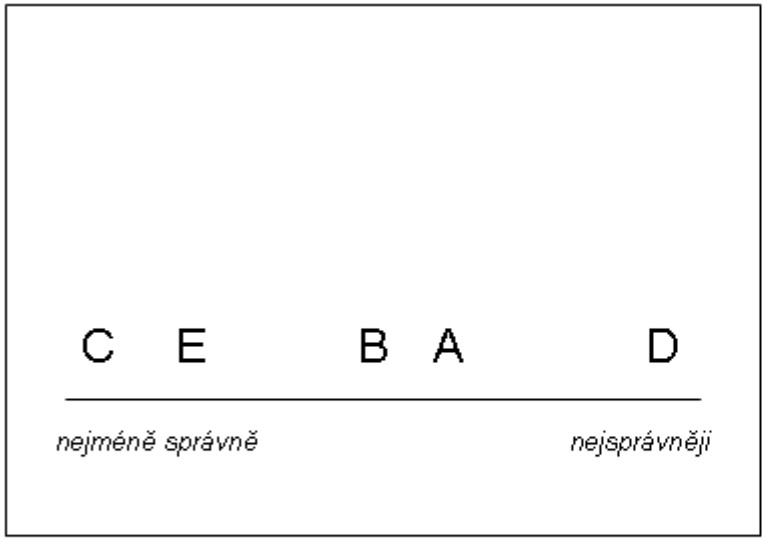
1. Ptejte se na **významné** problémy
   - Triviální či naopak příliš složité otázky vás jen připraví o čas
   - Vyvarujte se „chytáků“
2. Testujte **využití** znalostí, nikoliv izolovaná fakta
   - Otázka bývá dlouhá, odpovědi krátké
   - Otázku zpracujte jako „medailonek“

Pět zásad

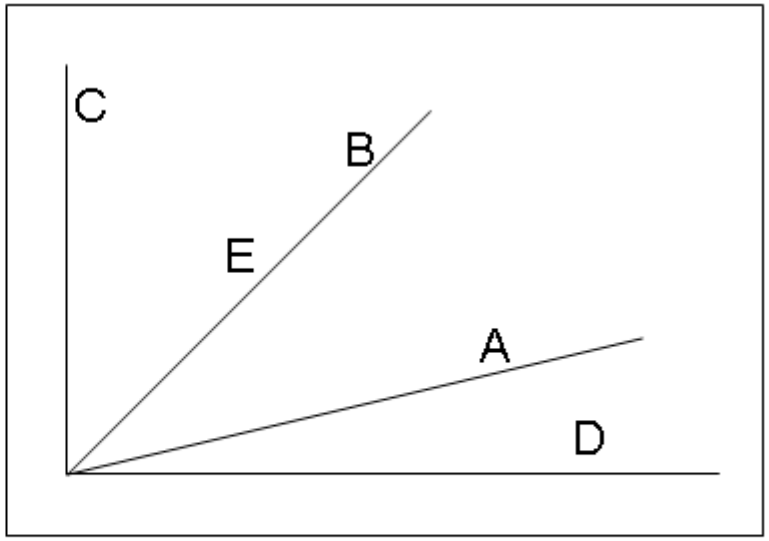
3. Formulujte **jasně a jednoznačně**
   – Odborník správně odpoví i se **zakrytými odpověďmi**
4. Pozor na slova „vždy“, „většinou“, „zřídka“, „výjimečně“, „nikdy“

Pět zásad

5. Nabídnuté odpovědi musí být **homogenní**



Správně:
nabídnuté možnosti jsou homogenní



Špatně:
nabídnuté možnosti spadají do různých kategorií

Otázka MTF:

Akutní intermitentní porfyrie je podmíněna poruchou biosyntézy

- ❑ **kolagenu**
- ❑ **kortikosteroidů**
- ❑ **mastných kyselin**
- ❑ **hemu**
- ❑ **tyroxinu**

Tvorba SBA

Dosud zdravý 33letý muž přichází pro epizody křečovitých bolestí břicha a svalové slabosti, které se vyskytují v posledním půl roce. Podobné obtíže mívá teta a bratranec. V průběhu epizody je břicho vzedmuté, peristaltika obleněná. V neurologickém nálezu je snížená síla velkých svalů horních končetin. Nález odpovídá defektu biosyntetické dráhy pro

(vyberte jedinou nejlepší odpověď)

- ❑ kolagen
- ❑ kortikosteroidy
- ❑ mastné kyseliny
- ❑ hem
- ❑ tyroxin

mefanet 2012

MEDICAL FACULTIES NETWORK

Standardizace testů

Cíl standardizace

- určení hranice pro úspěšné absolvování testu
- stanovení mezí pro zařazení do určité výkonnostní kategorie

samotné skóre nemá vypovídající hodnotu o tom, jak student v testu obstál v porovnání s ostatními

vyjádření výsledků jednotlivých respondentů vzhledem k výsledkům reprezentativního vzorku

Relativní vs. absolutní standardizace

Relativní standardizace

- Základem je normovaná metoda, která porovnává výsledky studentů mezi sebou
- Příkladem je percentilová škála

Absolutní standardizace

- Základem je kriteriální metoda, která vychází z počtu správných odpovědí jednotlivých studentů
- Příkladem je stanovení hranice 70% pro úspěšné složení testu

Relativní standardizace

- Nevztahuje se k obsahu testu, ale porovnává studenty mezi sebou
- Není nutné standardizovat každý test zvlášť
- Kolísání kvality úspěšných studentů podle kvality dané skupiny
  - Uspějí i slabší studenti, protože celá skupina byla slabší
  - V každém testu určitá část studentů neuspěje bez ohledu na znalosti
- Určuje se na základě dat získaných pilotováním nebo ostrým testováním
- Percentilová škála, Z-škála

Abolutní standardizace

- Eliminuje některé nevýhody relativní standardizace
  - (závislost na skupině)
- Lépe rozlišuje studenty, kteří mají lepší znalosti
- Základem je stanovení hranice mezi úspěšným a neúspěšným studentem
  - V praxi často pouze intuitivní přístup bez hlubšího zdůvodnění

→ Testy mohou být příliš jednoduché nebo naopak příliš obtížné

- Metody využívají expertní posudek odborníků na položky testu a testované studenti
  - Angoffova a Ebelova metoda

Angoffova metoda

- Minimálně kompetentní student (MKS)
  - Reprezentuje nejslabšího studenta, který by měl test zvládnout
- Tým pedagogů/expertů
  - Každý doplní jednotlivé otázky o počty MKS, kteří by měli otázky zodpovědět správně
  - Nézavislé hodnocení bez vzájemných konzultací
  - Jednotlivé hodnoty se poté dále zpracovávají

Angoffova metoda

Jaký světadíl označuje zelený kruh na vlajce olympijských her:

a) Austrálie
b) Amerika
c) Asie
d) Evropa

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Průměr
Otázka						
					Celkem	? = ? %

Angoffova metoda

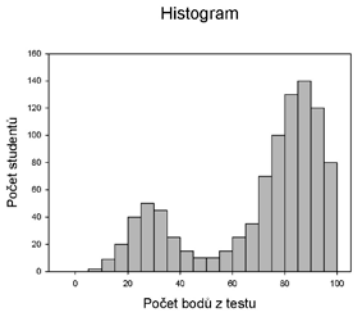
	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Průměr
Otázka 1	0,7	0,5	0,3	0,6	0,8	0,58
Otázka 2	0,8	0,6	0,5	0,7	0,4	0,6
Otázka 3	0,7	0,6	0,6	0,7	0,6	0,64
					Celkem	**0,61 = 61%**

mefanet 2012

MEDICAL FACULTIES NETWORK

Analýza výsledků testu

Proč analyzovat výsledky testu?

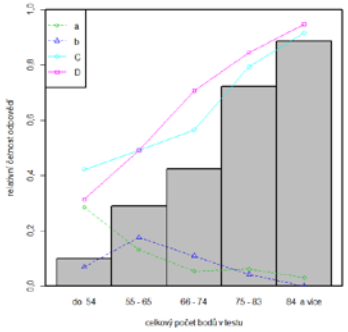


Liší se výsledky oproti loňsku?

Není test „vynesený“?

Měří test dostatečně přesně?

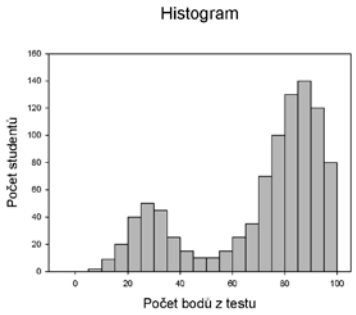
Měří test to, co chceme, aby měřil?



Jak kvalitní jsou jednotlivé položky?

Jak vhodné jsou nabízené distraktory?

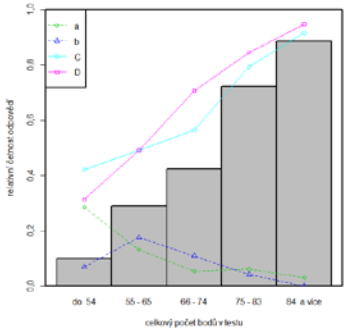
Analýza výsledku testu - OSNOVA



1. Popis celkového výsledku testu

2. Hodnocení kvality testu jako celku:

Reliabilita a validita



3. Hodnocení kvality položek

Jaká data analyzujeme a jak?

	úloha 1	úloha 2	úloha 3	úloha 4	úloha 5	CELKEM
Karel K.	15	15	0	10	1	41
Andrea K.	15	14	0	9	1	39
Iveta O.	15	14	1	10	1	41
Zdeněk T.	14	14	0	9	0	37
Ota L.	15	13	0	9	1	38
Helena J.	15	14	0	8	0	37
Ondřej K.	14	11	0	10	1	36
Helena P.	15	9	1	8	1	34
Jan T.	13	9	0	10	0	32
Iva T.	15	7	0	8	1	31
Anna A.	15	9	0	7	NA	31
Jan L.	15	10	0	6	0	31
Zdena S.	14	11	1	4	0	30
Jakub R.	15	6	1	7	1	30
Ondřej H.	15	7	1	5	0	28
Kateřina A.	15	7	0	5	0	27
Pavel L.	15	6	1	4	NA	26
Petr H.	15	7	1	NA	NA	23
Ctirad M.	15	6	1	NA	NA	22
Jan K.	13	0	1	NA	NA	14

Jaká data analyzujeme a jak?

ID	CH1	CH2	CH3	CH4	CH5	CH6	CH7	CH8	CH9	CH10	CH11	CH12	CH13	CH14	C
6551	CX	BX	AX	ABCDX	A	ABX	ABCX	AX	AC	DX	AX	CX	DX	BC	C
6552	CX	BX	AX	ABCDX	BX	ABX	ABCX	AX	CX	DX	AX	CX	DX	ADX	C
6553	CX	BX	C	ABCDX	A	ABX	ABCX	AX	CX	DX	AX	CX	DX	ACD	C
6554	CX	NULL	AX	B	BC	ABX	AB	AX	CX	B	NULL	B	B	ADX	C
6555	A	AB	CX	B	BC	A	AC	ABD	ABCD	ABD	C	BX	C	C	A
6556	CX	BX	AX	ABCDX	A	A	ABCX	AX	CX	DX	AX	D	DX	ADX	C
6557	CX	BX	AX	ABCDX	D	ABX	ABCX	AX	CX	DX	AX	B	DX	C	C
6558	CX	BX	AX	ABCDX	BX	ABX	C	AX	CX	DX	AX	CX	DX	C	C
6559	CX	A	C	ABCDX	D	ABX	ABCX	AX	A	DX	AX	CX	DX	ADX	C
6560	CX	A	AX	ABCDX	BX	ABX	AB	AX	CX	DX	AC	CX	DX	ADX	C
6561	CX	BX	AX	ABCDX	BX	ABX	ABCX	AX	CX	DX	AX	CX	DX	ADX	C
6562	CX	C	AX	ABCDX	BX	ABX	ABCX	AX	AC	DX	AX	CX	DX	ADX	C
6563	CX	BX	C	ABCDX	BX	ABX	AC	AX	D	DX	AX	CX	DX	A	C
6564	CX	A	D	C	D	ABX	ABCX	AX	CX	DX	AC	CX	B	ADX	C
6565	CX	BX	AX	ABCDX	BX	AD	AC	AX	CX	DX	B	B	DX	ADX	C
6566	CX	BX	AX	ABCDX	BX	A	ABCX	AX	CX	DX	AX	CX	DX	ADX	C
6567	CX	C	AX	C	BX	ABX	ABCX	AX	CX	DX	AX	A	DX	ADX	C
6568	C	BX	CX	A	CX	DX	BCX	BCDX	BCX	B	BX	BX	BC	AX	D
6569	BX	BX	CX	CX	CX	DX	C	BC	C	B	BX	BX	ACX	AX	D
6570	CX	BX	AX	ABCDX	BX	ABX	ABCX	AX	CX	DX	AX	CX	DX	ADX	C
6571	C	BX	CX	CX	CX	DX	BCX	BCDX	BCD	AD	BX	BX	ACX	AX	D
6572	CX	BX	AX	ABCDX	BX	ABX	ABCX	AX	CX	DX	AX	CX	DX	ADX	A
6573	CX	D	D	ABCDX	BX	B	ABCX	AX	CX	DX	AX	CX	DX	ADX	C
6574	CX	BX	AX	ABCDX	BX	ABX	ABCX	AX	CX	DX	AX	CX	DX	ADX	A
6575	CX	BX	AX	C	BC	ABX	ABCX	AX	AC	A	AX	CX	C	ADX	C
6576	CX	A	C	ABCDX	BX	ABX	ABCX	AX	CX	DX	AX	CX	DX	A	C
6577	CX	BX	AX	ABCDX	BX	ABX	ABCX	AX	CX	DX	AX	CX	DX	ADX	C
6578	CX	BX	AX	ABCDX	BX	ABX	ABCX	AX	CX	A	AC	CX	DX	ADX	C
6579	CX	BX	AX	ABCDX	BX	A	ABCX	AX	CX	DX	AX	A	DX	A	B
6580	CX	BX	AX	ABCDX	BX	ABX	AC	AX	CX	DX	AC	A	DX	ADX	C
6581	CX	BX	AX	ABCDX	BC	A	AC	AX	CX	DX	AX	CX	DX	ADX	C
6582	CX	BX	AX	ABCDX	BX	ABX	ABCX	AX	CX	DX	AX	CX	DX	ADX	C
6583	CX	A	AX	ABCDX	D	ABX	BCD	B	AC	DX	AC	CX	DX	A	A
6584	CX	BX	AX	ABCDX	BX	ABX	AB	AX	BC	DX	AX	CX	BCD	B	C
6585	CX	A	C	ABCDX	BX	ABX	ABCX	AX	D	A	AX	CX	DX	BC	B

Popis výsledků testu

Jsou výsledky očekávatelné?

Liší se různé testované skupiny?

Liší se úspěšnost v testu oproti loňsku?

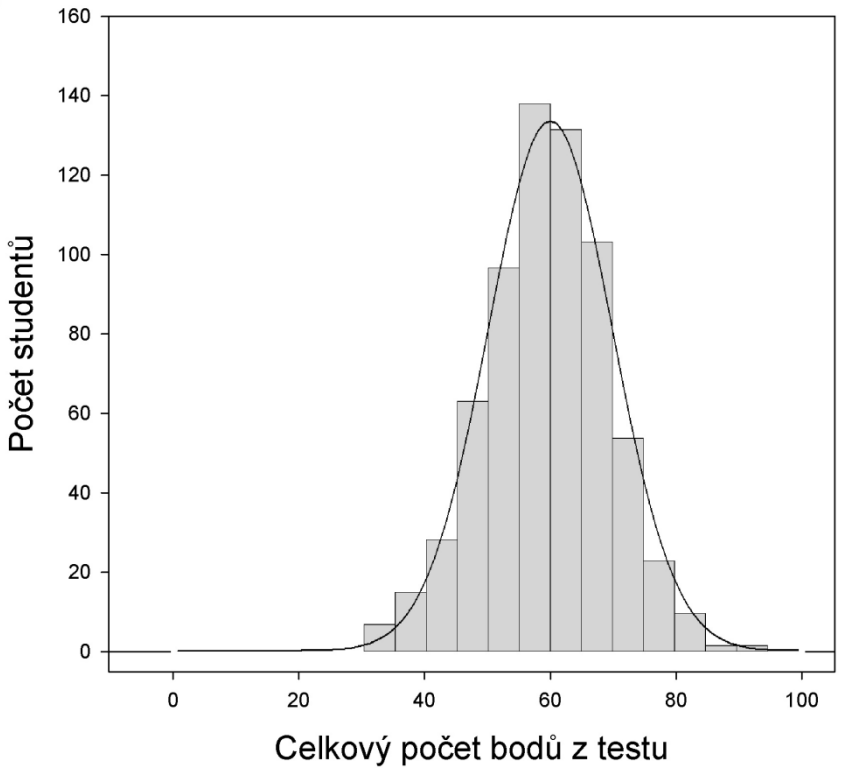
Popisné statistiky:

- průměr, medián, modus
- rozpětí, rozptyl, percentily

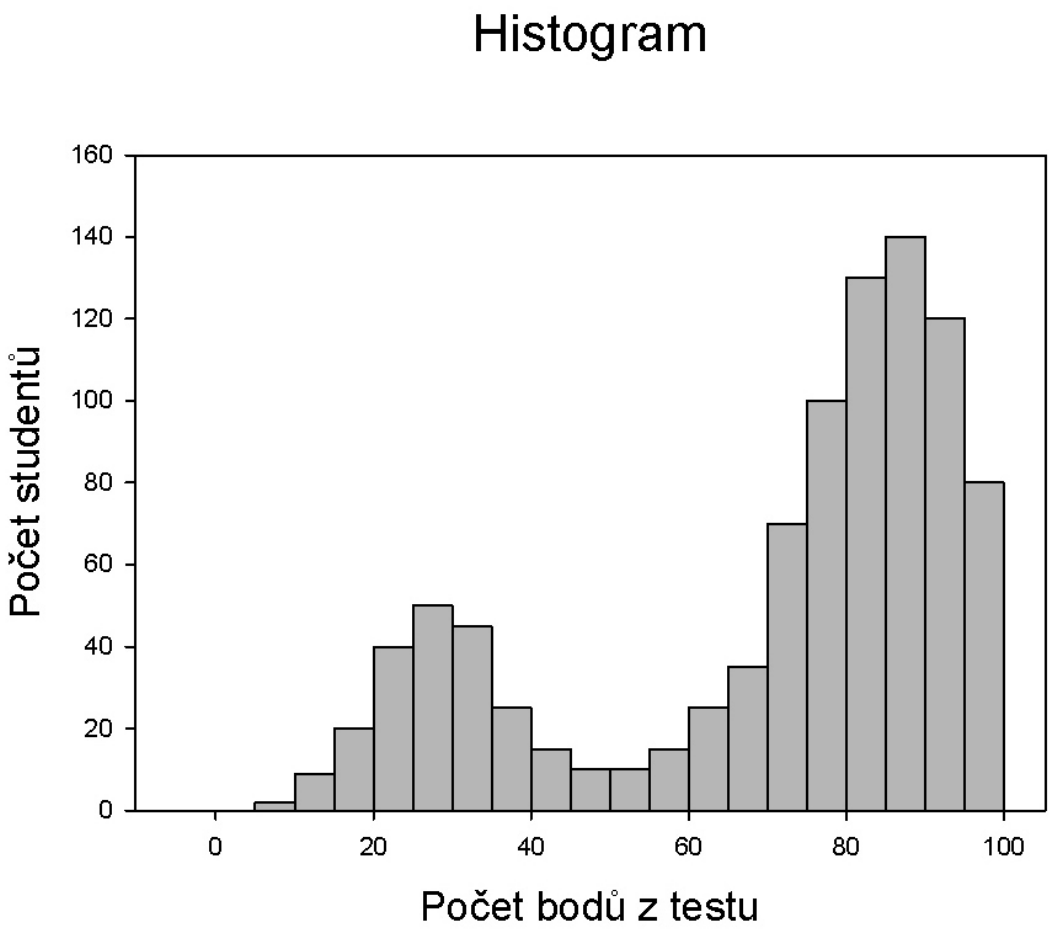
Graficky:

- histogram

Histogram



Podezřelé rozložení výsledků



Hodnocení kvality testu jako celku

Reliabilita

Měří test dostatečně přesně?

Jsou výsledky testu zopakovatelné?

Jak velký podíl variability přísluší *chybě měření*?

Validita

Měří test to, co chceme, aby měřil?

Reliabilita (spolehlivost) testu

Jsou výsledky testu zopakovatelné?

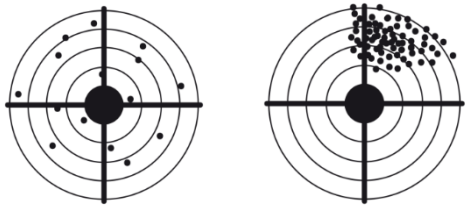
Zdroj: http://men-in-black-3-movie-trailer.blogspot.cz/

Test-retest reliabilita

Zopakování téhož testu stejnými studenty

Korelace mezi dvěma výsledky

Nevhodné: nadhodnocení při krátkém intervalu (zapamatování), efekt učení při delším intervalu



Reliabilita paralelních forem testu

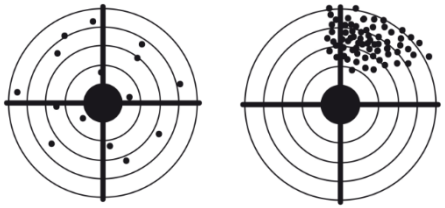
Zadání dvou „podobných“ verzí testu

Korelace mezi dvěma výsledky.

Náročné na tvorbu a administraci dvou verzí testu, únava žáků

Reliabilita (spolehlivost) testu

A máme-li pouze jediný test? Můžeme jej rozdělit!



Split-half reliabilita

Využívá korelace mezi dvěma částečnými výsledky

Které rozdělení na polovinu zvolit?

Cronbachovo alfa

Využívá korelace mezi jednotlivými položkami

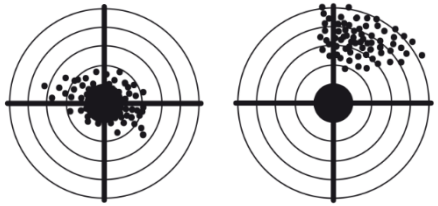
Mírou vnitřní konzistence testu

Vzorec implementovaný v každém statistickém softwaru

Ale bacha na Cronbacha!

Sijtsma K (2009): On the use, the misuse… of Cronbach's alpha. *Psychometrika.*

Validita testu



Měří test to, co chceme, aby měřil?

Obsahová validita: posouzení, zda test pokrývá zkoušenou látku

Kriteriální validita: vyžaduje další proměnnou (kritérium)

Využití korelačního koeficientu nebo regresní analýzy

Příklad: **Predikují** přijímací zkoušky úspěšnost studia?

Příklad: Koreluje výsledek testu s hodnotou v **souběžném** testu?

Příklad: **Přidává** test novou **informaci** nad již existující test?

Byčkovský, Zvára (2007): Konstrukce a analýza testů pro přijímací řízení.

Položková analýza

Jak je položka **obtížná**?

Je položka **citlivá**? Rozliší dobré a slabé studenty?

Je položka **spravedlivá**, měří všem stejně?

Proč byla položka tolikrát **vynechaná**?

Jak často a kým byly voleny jednotlivé **distraktory**?

Obtížnost položky

Relativní četnost správných odpovědí: $P = {}^{N_S}/_{N}$

Pro bodovanou položku normovaný průměr: $P = {}^{\bar{X}}/_{X_{max}}$

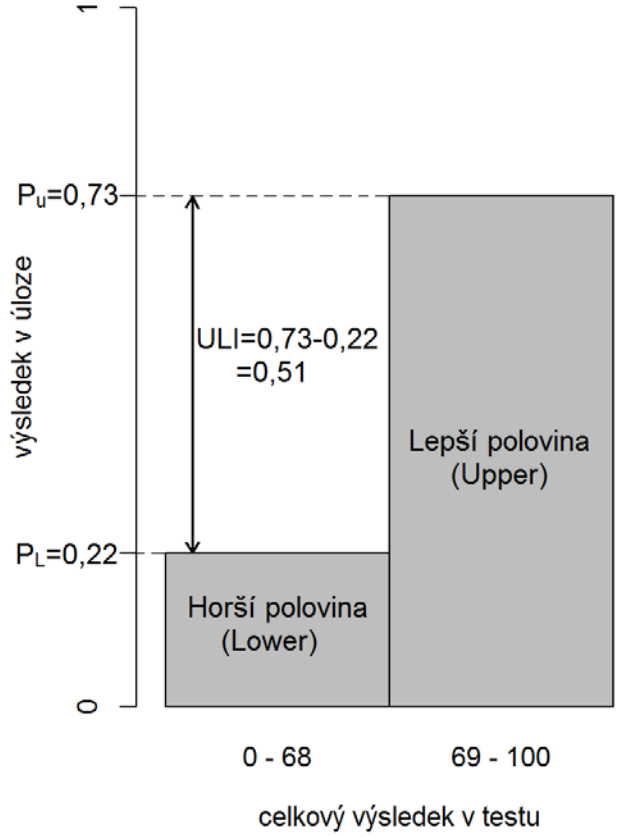
Velmi snadné položky na začátku testu - zvýšení motivace

Vzrůstající obtížnost položek v testu

Položky s obtížností okolo 0,5 mají nejlepší rozlišovací schopnost

Citlivost položky

Citlivost dle indexu ULI



- **Upper-Lower Index:**

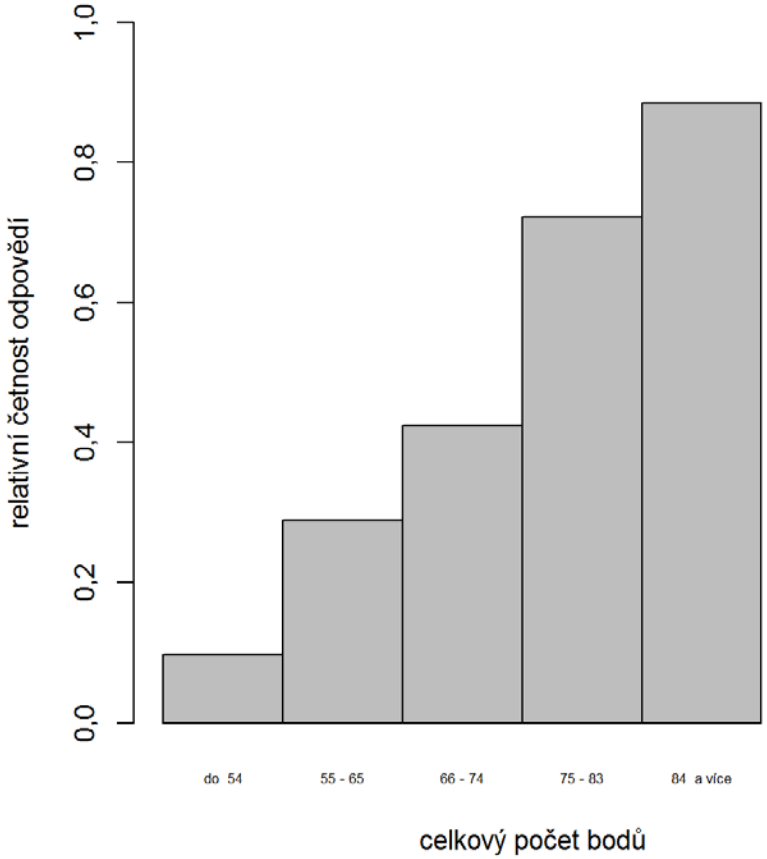
$$ULI = P_U - P_L$$

nebo

- **Korelace mezi položkovým skóre a celkovým počtem bodů**

- Vždy by měla být kladná!
- Čím je větší, tím lépe
- Hodnoty blízké 0 lze čekat jen u velmi snadných či velmi obtížných položek

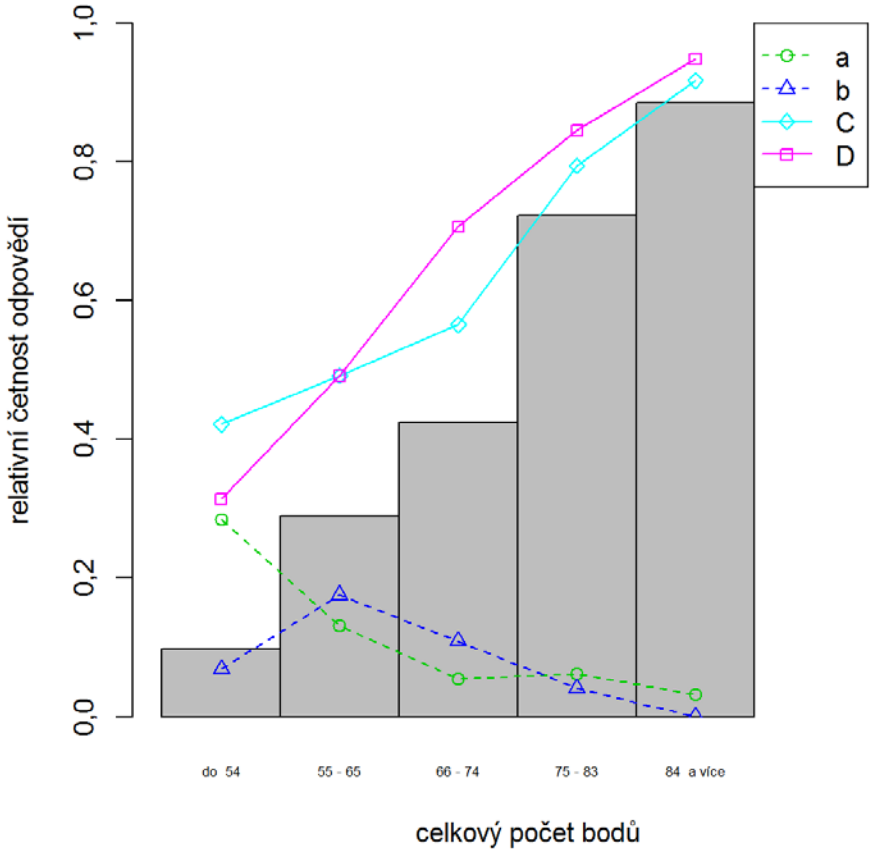
Detailní zobrazení vlastností položky



VHODNÁ POLOŽKA:

- Lepší žáci volí správnou odpověď častěji
- Rostoucí tendence
- Velký **sklon je znakem citlivosti**, tedy schopnosti rozlišit mezi lepšími a slabšími studenty

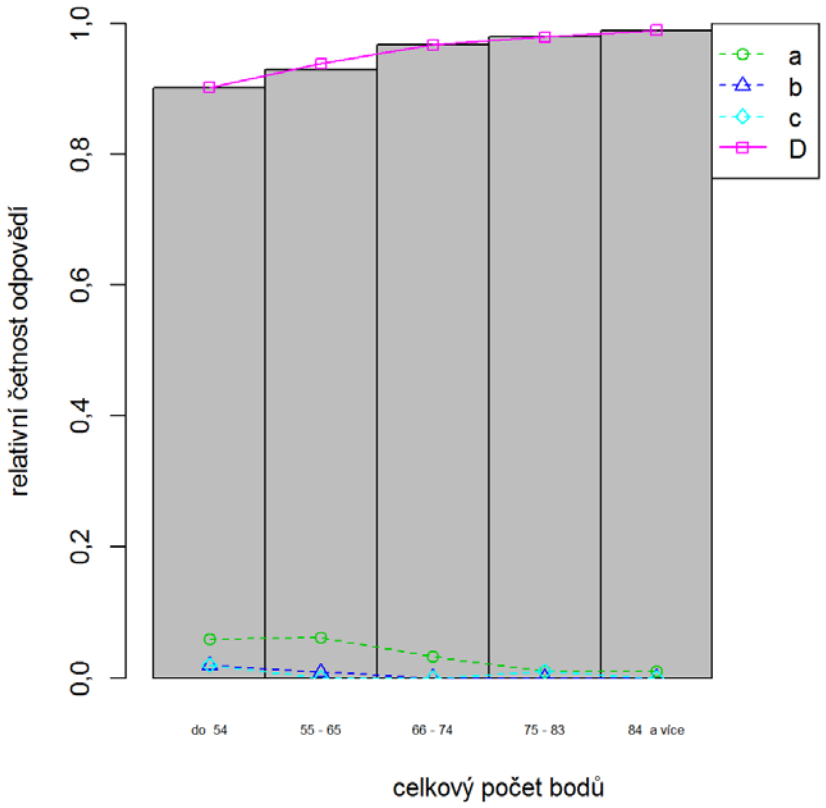
Detailní zobrazení vlastností položky



VHODNÁ POLOŽKA:

- Lepší žáci volí správnou odpověď častěji
- Rostoucí tendence
- Velký **sklon je znakem citlivosti**, tedy schopnosti rozlišit mezi lepšími a slabšími studenty
- Klesající tendence distraktorů (nesprávných odpovědí)

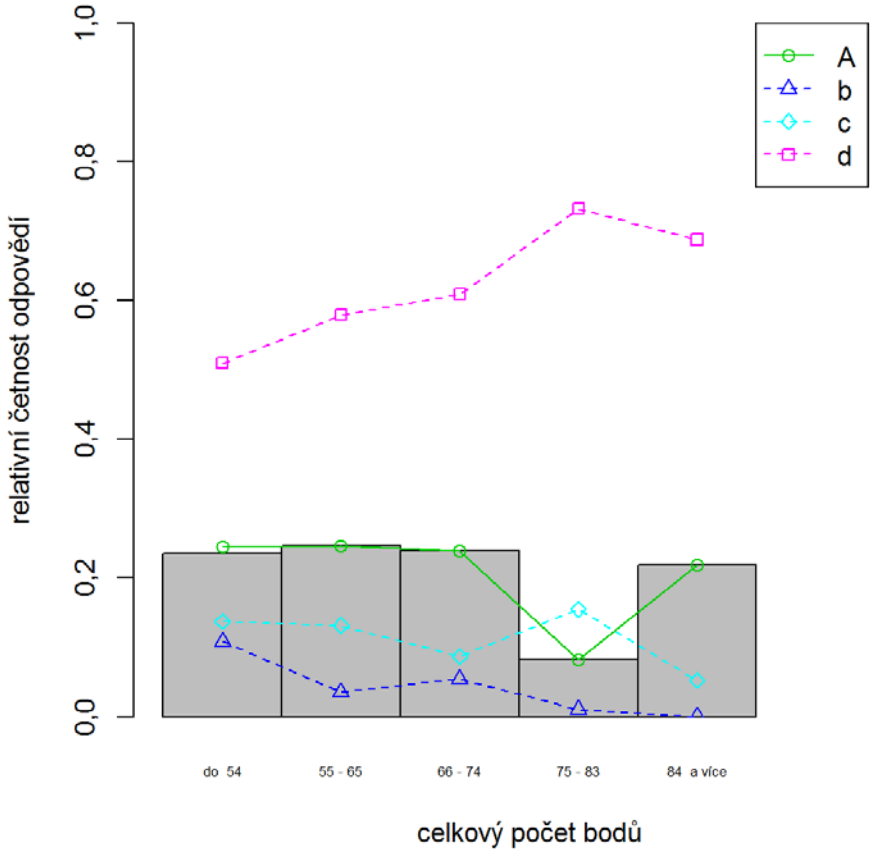
Detailní zobrazení vlastností položky



SNADNÁ POLOŽKA

- Distraktory b, c jsou zcela neatraktivní
- Položka špatně rozliší mezi lepšími a horšími studenty (malá diskriminační schopnost položky)

Detailní zobrazení vlastností položky



NEVHODNÁ POLOŽKA:

- Správná odpověď A volena málo a nezávisle na celkovém výsledku v testu
- Nejvíc studentů volí distraktor d
- Distraktor d volí častěji celkově lepší studenti
- Není distraktor d nejblíž správné odpovědi?
- Není položka nesrozumitelná?

Kdy klasické odhady nestačí?

Odhady **obtížnosti položky** se budou lišit

- zadáme-li test v prvním ročníku
- zadáme-li test ve třetím ročníku

Odhad obtížnosti položky je závislý na úrovni znalosti studentů!

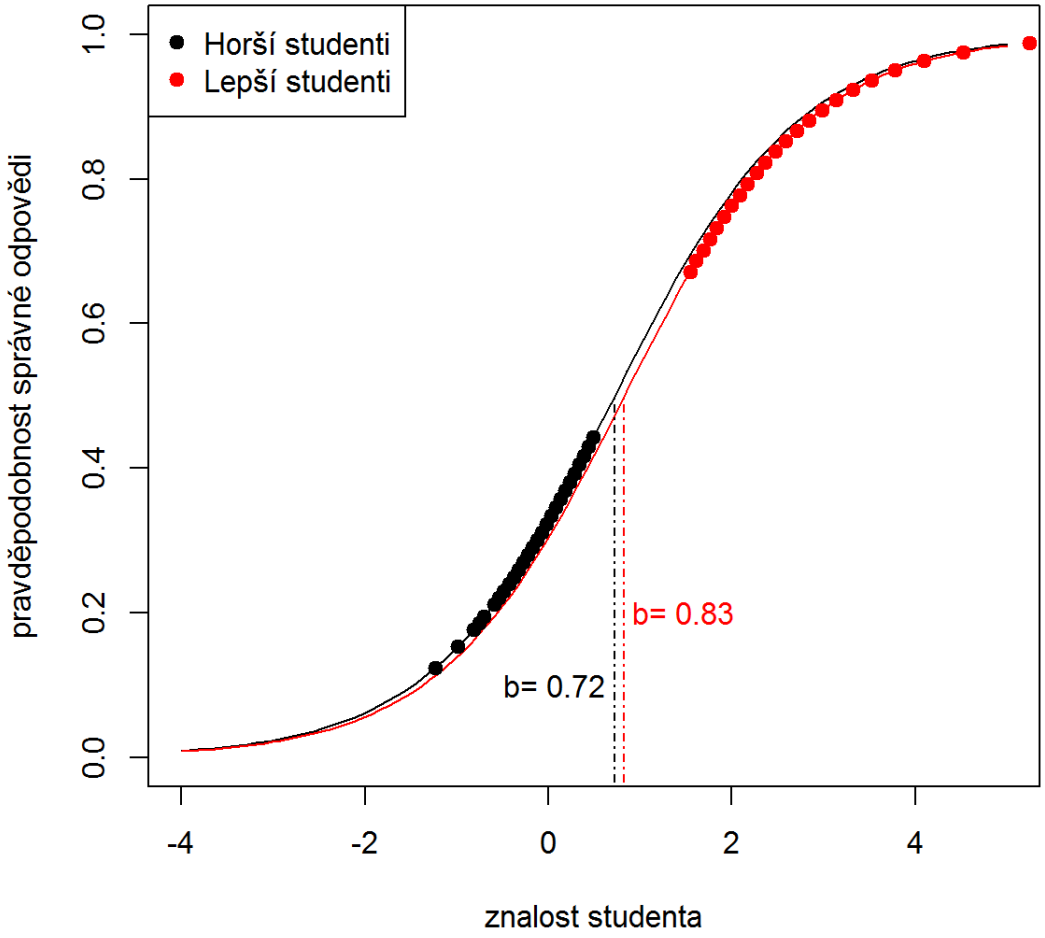
Odhady **citlivosti položky** se budou lišit

- zadáme-li test studentům z jednoho ročníku
- zadáme-li test studentům celé fakulty

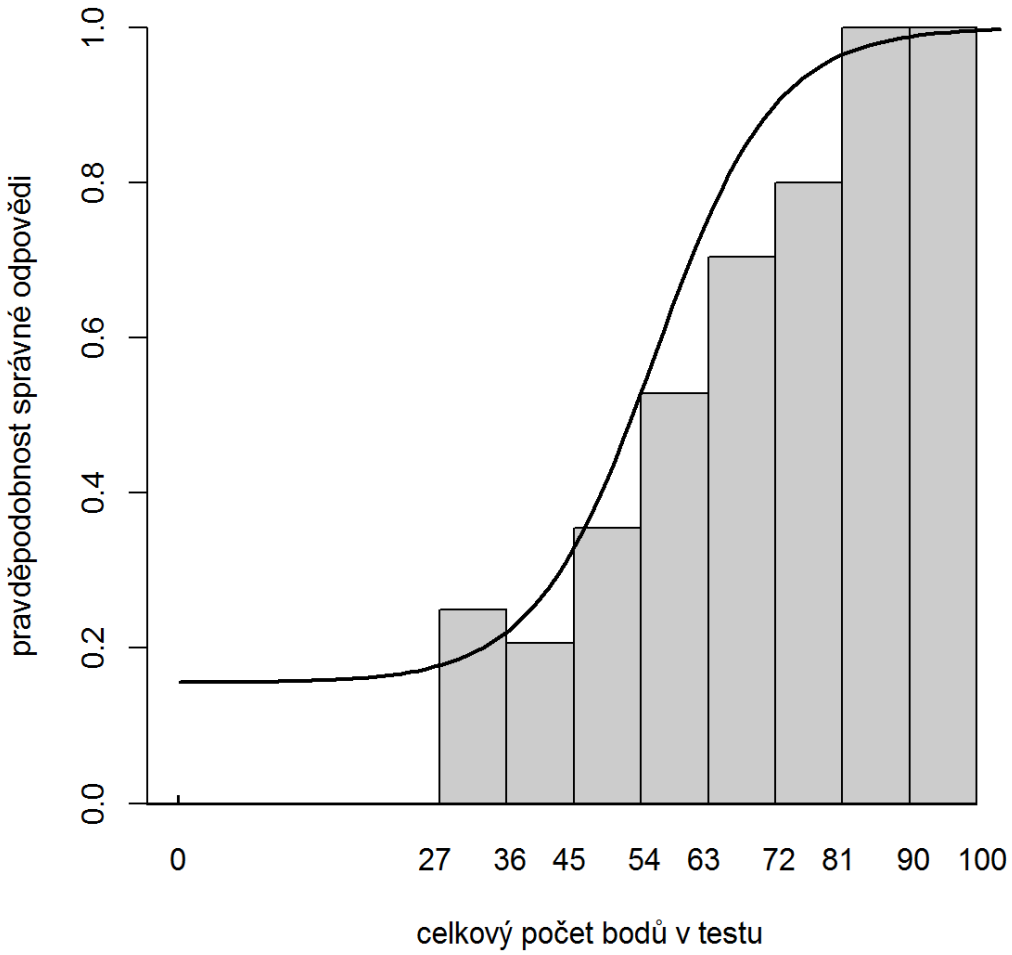
Odhad citlivosti je závislý na homogenitě testovaných studentů!

Pokud položku zadáváme různým skupinám, hodí se odhady zavést tak, aby byly **nezávislé na celkové úrovni znalosti** studentů

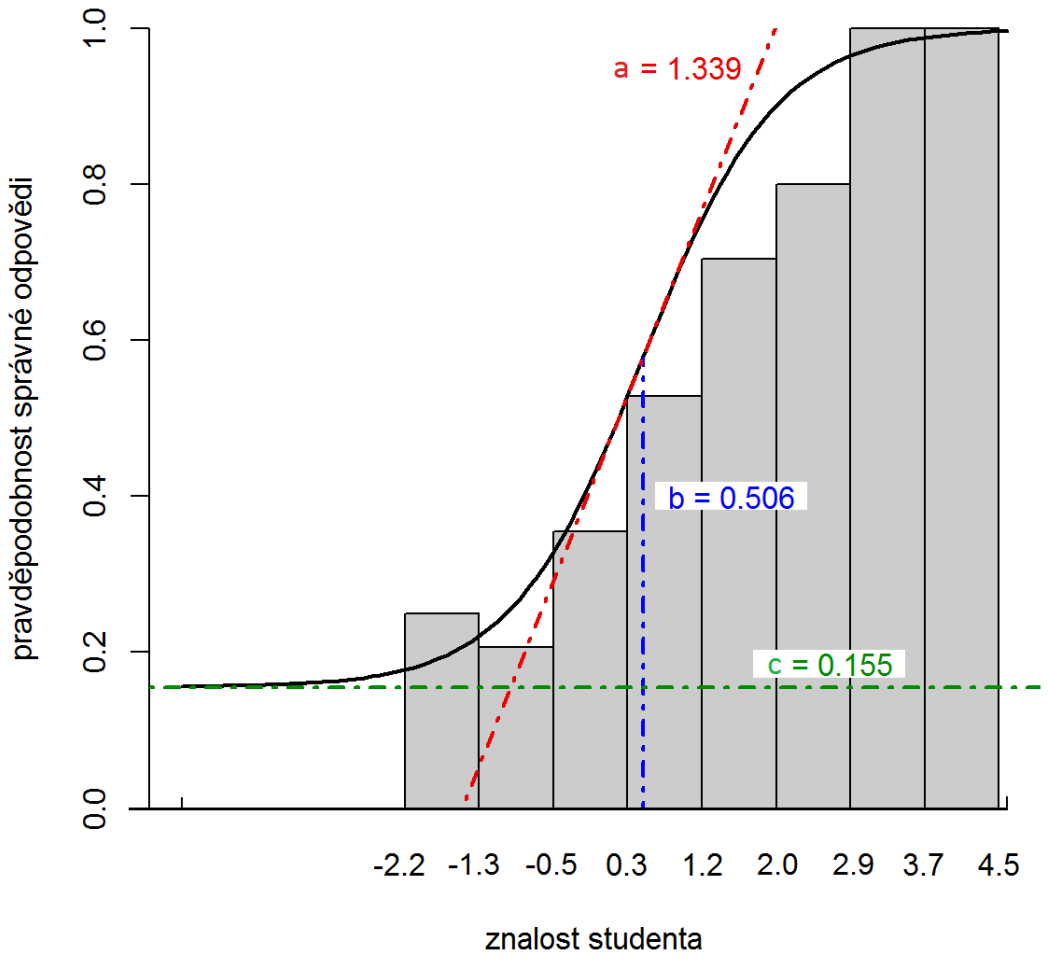
Teorie odpovědi na položku (IRT)



Teorie odpovědi na položku (IRT)



IRT odhady vlastností položky



b - obtížnost

a - citlivost

c - uhádnutelnost

- Vyžadují velké množství (stovky, ještě lépe tisíce) studentů!

- Vhodné pro velké databáze otázek (MCAT, SAT, TOEFL,...)

Analýza kvality testu - SHRNUTÍ

1. **Kontrola výsledků testované skupiny**, zda jsou očekávatelné
   - není test vynesený?
2. **Analýza kvality testu jako celku**
   - reliabilita, tedy spolehlivost testu
   - validita, tedy zda test měří to, co zamýšlíme
3. **Analýza položek testu**
   - odhady **obtížnosti** a **citlivosti**
   - IRT odhady vhodné pro větší databáze

-> přeformulování nebo vynechání nevhodných otázek

-> úprava počtu položek

-> případně i změna složení testu

mefanet 2012 | MEDICAL FACULTIES NETWORK

Moderní trendy

Týmová práce

Sdílení otázek

Oponentura testů

Otázky s jedinou nejlepší odpovědí (SBA)