



MASARYKOVA UNIVERZITA

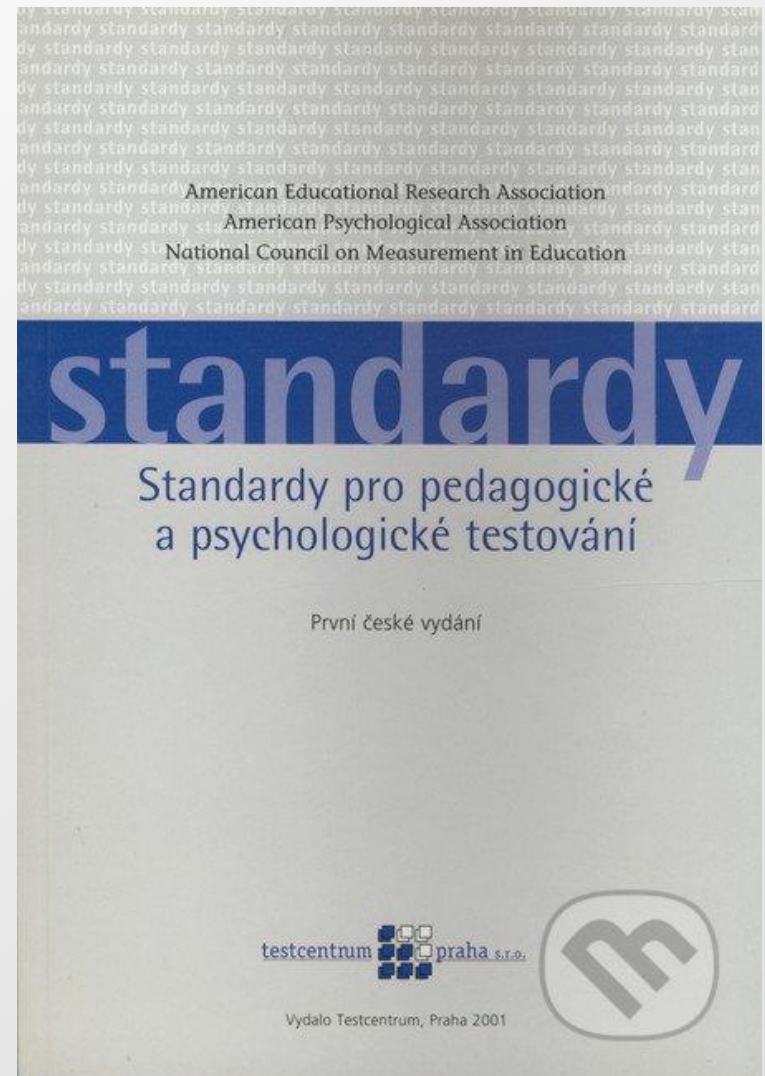
Konstrukce a analýza didaktických testů

Obsah příspěvku

- testy znalostí
- postup vytváření testu
- analýza vlastností položek a testu
- adaptivní testování

Zdroje poznatků

- pedagogika, didaktika
- psychometrika
- etika (např. Standardy APA)



Testy znalostí

- specifika testů znalostí
- typy testů znalostí
 - podle měřeného výkonu
 - podle interpretace výsledku
 - podle účelu při hodnocení předmětu
 - podle míry standardizace

Postup konstrukce testu

- koncepce testu
 - vymezení univerza znalostí
 - formát položek, délka testu
 - postup administrace, vyhodnocení
- vytváření položkové banky
 - návrh a tvorba položek
 - oponentura, pilotáž
- sestavení testu (příp. paralelních forem)
- administrace testu a vyhodnocení

Koncepce testu

- Než začneme vytvářet položky:
 - jaké znalosti, dovednosti či schopnosti chceme testem zjišťovat?
 - jaká je role testu? (screeningový vstupní, zápočtový, zkouškový – s odstupňovaným hodnocením)
 - forma testu, časový limit, formát položek...

Konstrukce testu - základní pojmy

☞ Položka

- ☞ Zadání, správná odpověď (odpovědi), distraktory

☞ Položková banka

- ☞ Souhrn všech položek obsahově pokrývajících určitou tematickou oblast

☞ Test

- ☞ Sestava položek, vybraných z položkových bank za účelem testování konkrétní osoby na základě předem definovaného algoritmu

Konstrukce položkových bank a testů

- Vymezení univerza znalostí – náplň oboru definovaného např. sylabem vyučovaného předmětu
- Test musí být reprezentativní vzhledem k univerzu znalostí v daném oboru

Konstrukce testu – dosažení reprezentativnosti

- Rozdělení univerza znalostí na subdomény
 - Při procesu vytváření položek slouží k zajištění toho, aby položkové banky pokrývaly všechny oblasti, které má uchazeč ovládat
 - Při procesu administrace testu slouží k tomu, aby konkrétní test obsahoval položky ze všech těchto oblastí

Konstrukce testu – dosažení reprezentativnosti

- Vytvoření položkových bank dostatečně pokrývajících všechny aspekty daných subdomén
 - Posouzení navrhovaných položek více experty ze dvou obecných hledisek (zajištění obsahové validity)
 - Kvalita navržené položky – jednoznačnost, správnost, tematická relevantnost
 - Kvalita celé položkové banky – určení nepokrytých nebo nedostatečně pokrytých témat problematiky

Konstrukce testu – dosažení reprezentativnosti

- Navržení vhodného algoritmu pro výběr položek z položkových bank do výsledných testů
 - Počet položek z každé položkové banky vybraných do konkrétního testu by měl odpovídat důležitosti dané problematiky pokryté položkovou bankou
 - Klasický náhodný výběr a jeho varianty
 - Doporučujeme zavést faktor obtížnosti položek a zohlednit ho při výběru vedle obsahové stránky
 - Pozice položky v testu – možný efekt praxe a efekt únavy
 - Adaptivní administrace testu
 - Moderní, stále více používaný přístup, založený na Teorii odpovědi na položku, nicméně stále technicky obtížně realizovatelný v našich podmínkách (počítačové adaptivní testování - CAT)

IS: Sady otázek v odpovědníku

Výběr sad otázek:

použit: zaškrtněte u sad, ze kterých se mají použít otázky

počet otázek: lze vybrat jen několik otázek (prvních nebo náhodně)

pořadí sady: jak mají jít sady za sebou - nevyplníte-li, jdou jak jsou zavedeny nebo náhodně

test_anova1 odp/tb/test_anova1.qdef			
<input checked="" type="checkbox"/> použít	<input type="checkbox"/> otázek z celkem 1 (nevyplněno = použít vše)	<input type="checkbox"/> pořadí sady	upravit sadu
test_freq2 odp/tb/test_freq2.qdef			
<input checked="" type="checkbox"/> použít	<input type="checkbox"/> otázek z celkem 1 (nevyplněno = použít vše)	<input type="checkbox"/> pořadí sady	upravit sadu
test_chisq1 odp/tb/test_chisq1.qdef			
<input checked="" type="checkbox"/> použít	<input type="checkbox"/> otázek z celkem 1 (nevyplněno = použít vše)	<input type="checkbox"/> pořadí sady	upravit sadu
test_metoda2 odp/tb/test_metoda2.qdef			
<input checked="" type="checkbox"/> použít	<input type="checkbox"/> otázek z celkem 2 (nevyplněno = použít vše)	<input type="checkbox"/> pořadí sady	upravit sadu
test_oneyway2 odp/tb/test_oneyway2.qdef			
<input checked="" type="checkbox"/> použít	<input type="checkbox"/> otázek z celkem 1 (nevyplněno = použít vše)	<input type="checkbox"/> pořadí sady	upravit sadu
test_ra2 odp/tb/test_ra2.qdef			
<input checked="" type="checkbox"/> použít	<input type="checkbox"/> otázek z celkem 1 (nevyplněno = použít vše)	<input type="checkbox"/> pořadí sady	upravit sadu
test_teor1 odp/tb/test_teor1.qdef			
<input checked="" type="checkbox"/> použít	<input type="checkbox"/> otázek z celkem 2 (nevyplněno = použít vše)	<input type="checkbox"/> pořadí sady	upravit sadu
test_ttest1 odp/tb/test_ttest1.qdef			
<input checked="" type="checkbox"/> použít	<input type="checkbox"/> otázek z celkem 1 (nevyplněno = použít vše)	<input type="checkbox"/> pořadí sady	upravit sadu

Vybrat ze všech sad otázek ze složky Odpovědníky Studijních materiálů tohoto předmětu

ll_el_ucitel?rakulta=1421;obdobi=5964;kod=PSA_032;predmet=749015;zuv=508744;akce=edit;qdescuri=%2F%2F1421%2FJano2014%2FPSA_032%2Fodp%2Ftb%2Ftes

IS: Pořadí otázek v testu

Pořadí otázek v rámci odpovědníku má být:

- zcela náhodné
- postupně otázky z jednotlivých sad, v rámci každé sady zamíchat náhodně
- zachovat posloupnost otázek tak, jak jsou za sebou v sadě

Pořadí odpovědí (jsou-li nějaké odpovědi nabízeny) má být:

- náhodné
- zachovat pořadí tak, jak je v definičních souborech

Typy položek a jejich konstrukce

- Eseje, překlady delších textů...
 - v ČR méně používaný typ
 - v zahraničí často významná součást přijímacích i jiných zkoušek
 - náročné na hodnocení a jeho objektivitu

- Otevřené položky s rozsáhlou odpovědí
 - snadná konstrukce x nesnadné objektivní hodnocení
 - někdy naznačená struktura odpovědi
 - často vhodnější ústní zkoušení

Typy položek a jejich konstrukce

- Otevřené položky s jednoznačnou odpovědí
 - Vhodné, pokud existuje jediná správná odpověď s omezeným spektrem formulace odpovědi
 - Odbourání faktoru uhádnutelnosti

- Příklady: doplnění nejvhodnějšího slova do textu v jazykovém testu, výsledek výpočtu v matematických příkladech

Typy položek a jejich konstrukce

➤ Dichotomní položky

- Ano / Ne
- Obrovský problém uhádnutelnosti, jediným řešením je penalizace
- Spíše nepoužívat

➤ Příklad:

- Rozhodněte a napište (P nebo N), zda-li je výrok pravdivý (P) či nepravdivý (N):
- Nejrozšířenější náboženství na Filipínách je římskokatolické.

Typy položek a jejich konstrukce

- ☒ Vícenásobná volba s jednou správnou odpovědí (radiobutton)
 - ☒ Nejčastěji používaný typ položky v testování znalostí
 - ☒ Problém uhádnutelnosti – řeší se množstvím kvalitně navržených distraktorů, případně penalizací nesprávné odpovědi

Příklad

S termínem noogenní neuróza se můžeme setkat

- a) v logoterapii V. E. Frankla
- b) v psychoanalýze S. Freuda
- c) v analytické psychologii C. G. Junga
- d) v teorii pole K. Lewina

Typy položek a jejich konstrukce

- ☒ Vícenásobná volba s různým množstvím správných odpovědí (checkbox)
 - ☒ nízká uhodnutelnost
 - ☒ problematický formát z hlediska skórování
 - ☒ Bod za správně určenou kombinaci odpovědí, nula v každém jiném případě
 - ☒ Nebo např. 1 bod za správně označenou možnost a -1 za nesprávně
 - ☒ Doporučujeme bodování 1 / 0 za celou položku, přičemž by mělo být možné při znalosti principu odvodit správnou kombinaci odpovědí a položka by neměla být formulačně komplikovaná

Typy položek a jejich konstrukce

➤ Přiřazovací položky

- vytvořit dvojice, seřadit možnosti (např. chronologicky)
- opět problematické skórování
- nižší uhádnutelnost

➤ Příklad: Utvořte správné dvojice:

- | | |
|-----------------------|-------------------|
| ➤ 1 ropa | A Doněcká pánev |
| ➤ 2 železná ruda | B poloostrov Kola |
| ➤ 3 fosfáty a apatity | C Povolží |
| ➤ 4 černé uhlí | D Bělorusko |

Obecná doporučení pro vytváření položek

- ☒ Zaměřte se na položky s jednou správnou odpovědí
 - ☒ možnosti odpovědí mohou být na kontinuu definovaném správností nebo vhodností odpovědi (s výrazným odstupem mezi správnou odpovědí a distraktory)
 - ☒ používejte dostatečný počet distraktorů (ideálně 3-5)
 - ☒ používejte formulaci „která odpověď nejlépe vystihuje...“
 - ☒ eliminujte tipování správné odpovědi odečtem části bodu za nesprávnou odpověď (podle počtu distraktorů)
 - ☒ pravděpodobnost uhádnutí správného řešení snížíme i zařazením neurčité odpovědi („žádná z nabízených není správná“)

Rady a doporučení – vytváření položek

- ❏ U položek s různým množstvím správných odpovědí (checkbox)
 - ❏ pokud nespecifikujeme počet správných odpovědí, musí být každá možnost jednoznačně správná nebo jednoznačně chybná (v praxi obtížně proveditelné)
 - ❏ možné je specifikovat v zadání počet odpovědí, které mají být identifikovány jako správné
- ❏ Vyhýbejte se záporným formulacím nebo dokonce dvojitým záporům
- ❏ Vyhýbejte se složitým větám v odpovědích – zadání může být dlouhé, varianty odpovědi ale krátké, stručné, jednoznačné
- ❏ Pozor na nezamýšlené nápovědi

Oponentura položek

- ☞ posouzení položek nezávislými odborníky
 - ☞ odstranění chyb v položkách, nejednoznačných odpovědí, nejasných formulací
 - ☞ posouzení vhodnosti obsahu položky
 - ☞ posouzení obtížnosti položky
 - ☞ posouzení distraktorů

doporučujeme použití strukturovaného hodnocení položek

Konstrukce testu – sestavení testu

- podle požadovaného počtu položek / časové délky testu
- rozdělení podle subdomén

- v případě potřeby vytvoření paralelních forem
 - po jednotlivých položkách/doménách – srovnatelnost obsahu
 - současně dbát na srovnatelnou obtížnost a časovou náročnost položek

Pilotáž testu / pretest

- často obtížně proveditelné
- snadnější otestovat jednotlivé položky než test jako celek
 - požadavek na utajení
 - pilotování položek formou procvičovacího testu - jiné odpověďové strategie než při testech
 - stanovení časového limitu

Konstrukce testu – skórování

- ☒ skórování jednotlivých položek
 - ☒ nejčastěji 0/1 a prostý součet
 - ☒ někdy podle obtížnosti či časové náročnosti položek různé váhy

- ☒ hodnocení výsledku testu
 - ☒ relativní vs. absolutní výkon
 - ☒ stanovení bodové hranice pro úspěšné zvládnutí učiva (či jeho subdomén)
 - ☒ stanovení bodových hranic pro jednotlivé stupně hodnocení

Analýza vlastností testu

- objektivita
- spolehlivost (reliabilita)
- platnost (validita)
- citlivost (rozlišovací účinnost)

Analýza vlastností testu

☞ Objektivita

- ☞ nezávislost výsledků testování na osobě administrátora testu, místě a čase testování atd.
- ☞ dosahujeme standardizací instrukcí, přiblížením podmínek testování jednotlivých studentů
- ☞ u hodnocení otevřených otázek ideálně více hodnotitelů

Analýza vlastností testu

☞ Reliabilita

☞ Vyjadřuje přesnost měření

☞ Relevantní metody zjišťování reliability

☞ Reliabilita paralelních forem / split-half

☞ Vyjadřuje míru shody mezi výsledky studentů se stejnou úrovní schopnosti, kterým jsou předloženy srovnatelné formy testu (či hypotetické poloviny).

☞ Test – Retest

☞ Míra shody mezi výsledky opakovaného testování stejných osob

☞ Vnitřní konzistence testu

☞ Užívá se v případě testů jednodimenzionálních

Reliabilita

- požadovaná minimální hodnota záleží na účelu testu
- odhad reliability ovlivněn např. počtem položek, heterogenitou souboru testovaných osob, rychlostní složkou v testu, obsahem i formou položek

IS: Export odpovědí pro analýzu vnitřní konzistence

Tisk a rozpoznání písemek Další

- Sestavit tištěné písemky (s výběrem odpovědníku)
- Tisk: Zadání písemek (s výběrem odpovědníku)
 - Odpovědní list
 - Průvodka (s výběrem odpovědníku)
- Rozpoznání obsahu složky
- Smazat ze složky záznamy
- Výstup do XML (s výběrem odpovědníku)
- Export pro dril
- Vyhledat duplicity
- Činnost během testu (s výběrem odpovědníku)
- Export odpovědí (s výběrem odpovědníku)

IS: Export odpovědí pro analýzu vnitřní konzistence

Odpovědi se exportují ve formě tabulky. První řádek obsahuje zadání otázky, na druhém jsou číselované jednotlivé podotázky (pokud je otázka složená z n odpovědí, vypisují se sloupce číselované 1..n a každý sloupec obsahuje odpověď na danou podotázku). Na dalších řádcích jsou jednotlivé odpovědi respondentů.

Nastavení exportu

Zadání otázek:

- textově
- číselně

Možnosti otázek:

- textově
- číselně

Průchody odpovědníku:

- pouze poslední uložení
- všechna uložení

Zobrazit otázky:

- pouze použité
- všechny otázky ze sad

Exportovat:

- pouze nejnovější podobu otázek ze sady otázek
- všechny předchozí verze upravených otázek

Odpovědi:

- pouze vybraných studentů
- všechny odpovědi

Vyberte typ výstupu:

- Formát pro Excel a Word (oddělovačem je tabulátor)
- Formát CSV pro Excel
- Textový tvar se sloupci oddělenými dvojtečkami
- Kontrolní výpis pouze na obrazovku
- Výstup pro tisk z prohlížeče

Analýza vlastností testu

➤ Validita testu

- Nakolik skór v testu vyjadřuje skutečnou úroveň znalostí studenta v dané oblasti

➤ Relevantní metody zjišťování validity

➤ Obsahová validita

- Reprezentují položky celé uvažované spektrum znalostí?
- Založena na expertním posouzení více experty

➤ Kriteriaální validita

- Vyjadřuje míru shody výsledku testu s vnějším kritériem, např. výsledkem ve srovnatelném testu, celkovou akademickou úspěšností atd.

Psychometrická analýza položek

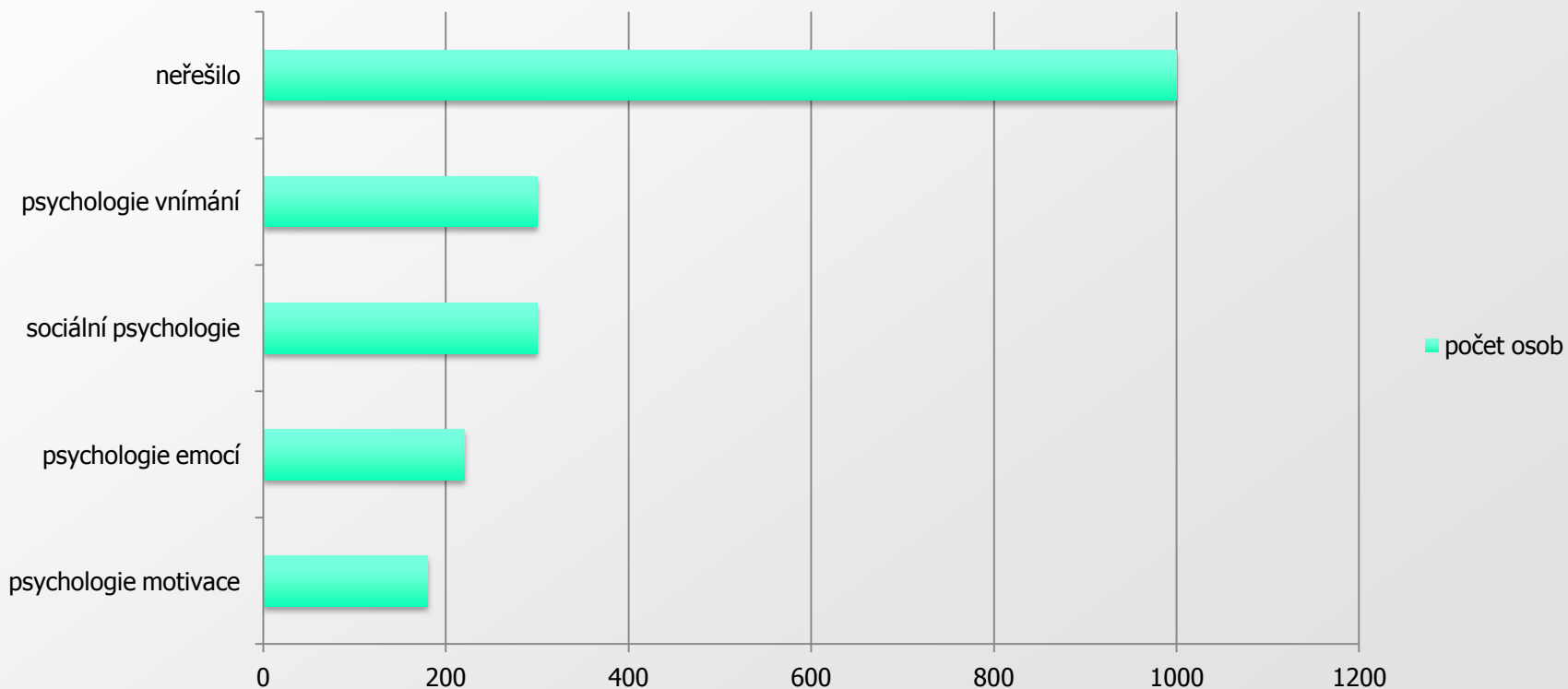
- ▣ obtížnost položky
- ▣ diskriminační (rozlišovací) účinnost položky

Psychometrická analýza položek

- Obtížnost vyjadřuje podíl osob, které položku vyřešily správně
- Ideální podíl mezi 20 – 80 %, současně analýza distraktorů
 - Jednoduchá položka ($p=0,95$):
 - Tato droga vyvolává halucinace již při nízkých dávkách. Halucinace mohou být v podobě živých barev a zvuků nebo i mystické povahy. O jakou drogu se jedná?
 - a) marihuana b) heroin c) **LSD** d) pervitin
 - Obtížná položka ($p=0,15$):
 - S Amesovou místností se můžeme setkat především ve výzkumech v oblasti:
 - a) psychologie motivace b) psychologie emocí
 - c) sociální psychologie d) **psychologie vnímání**

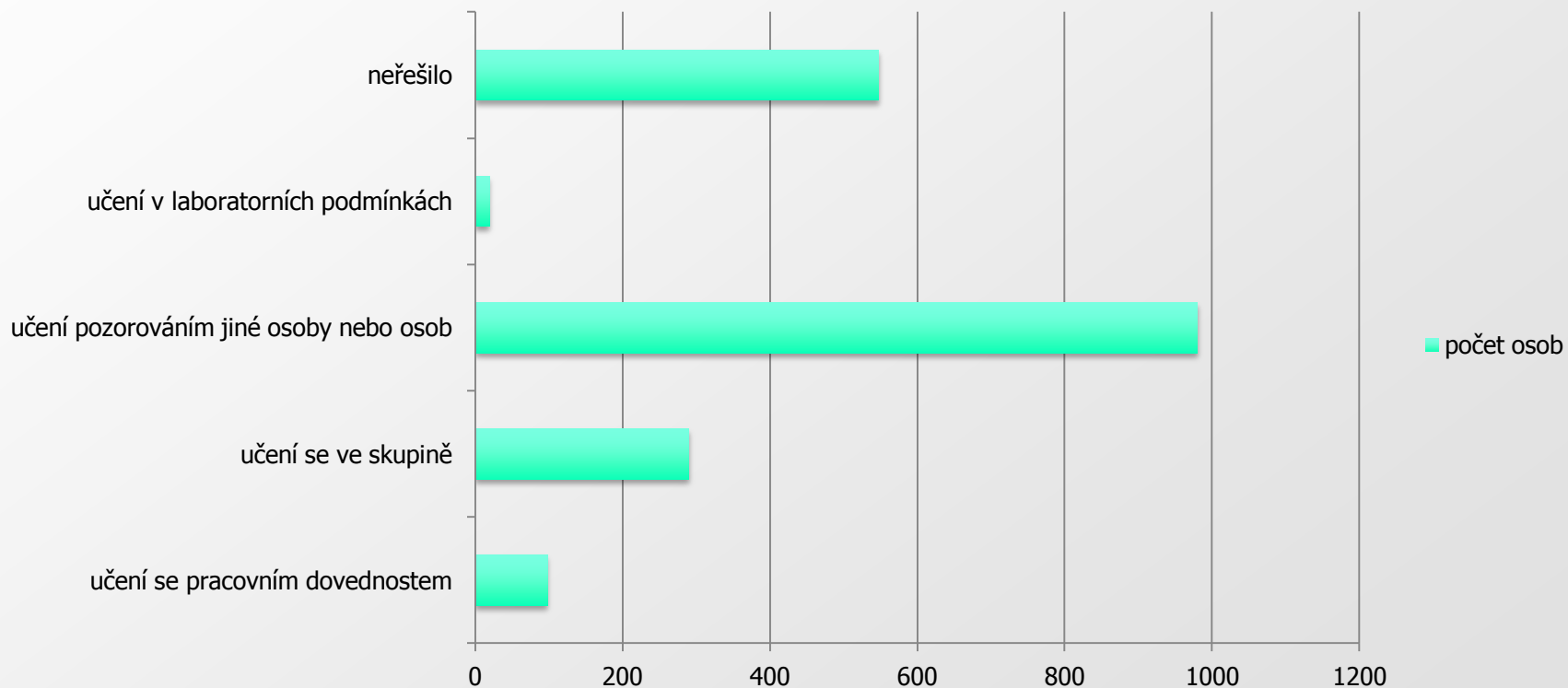
Analýza obtížnosti položky a distraktorů

Amesova místnost



Analýza obtížnosti položky a distraktorů

Observační učení



Analýza distraktorů při testování v IS

Počet odpovídajících: 21 / 21

- [zvolit](#): pro MM - zkouška 14/5 ve 12:00
/el/1421/jaro2014/PSA_038/odp/tb/zk02MM.qdesc
Počet odpovídajících: 1 / 1
- [zvolit](#): zkouška 14/5 ve 12:00
/el/1421/jaro2014/PSA_038/odp/tb/zk02.qdesc
Počet odpovídajících: 11 / 11
- [zvolit](#): zkouška 27/5 v 8:40
/el/1421/jaro2014/PSA_038/odp/tb/zk03.qdesc
Počet odpovídajících: 18 / 18

Sady otázek a odpovědníky

- [Sady otázek](#)
- [Odpovědníky](#)
- [Popisy odpovědníků](#)
- [Zodpovězení odpovědníku](#)
- [Zkontrolovat odpovědníky](#)
- [Přepočítat body u odpovědníku](#)

Správa odpovědí

- [Odpovědi](#)
- [Obtížnost](#)
- [Rušení odpovědí](#)
- [Obnovení zrušených odpovědí](#)
- [Korekce správných/špatných odpovědí v hodnotící části](#)

Tisk a rozpoznání písemek

- [Sestavit tištěné písemky](#)
- Tisk: [Zadání písemek](#)
[Odpovědní list](#)
[Průvodka](#)
- [Rozpoznání obsahu složky](#)
- [Smazat ze složky záznamy o rozpoznání](#)
- [Vyhodnocení rozpoznání písemek](#)
- [Nahlížení do odpovědních listů](#)

Analýza distraktorů při testování v IS

V recenzním posudku diplomové práce *Antecedenty a koreláty životní spokojenosti u adolescentů* se můžeme dočíst: "Zásadní omezení studie lze spatřit v použitém výzkumném nástroji na zjišťování životní spokojenosti. Použitý nástroj "Vobořilova Škála životní spokojenosti" podle výsledků nových studií neměří životní spokojenost, ale spíše aktuální emocionální ladění respondentů." Co recenzent kritizuje?

- externí validitu
- *operacionalizaci proměnných
- validitu statistických postupů
- reliabilitu nástrojů

<i>koeficient obtížnosti</i>	55
<i>počet použití otázky</i>	18
<i>počet správných odpovědí</i>	10
<i>počet chybných odpovědí</i>	7
<i>počet nezodpovězených</i>	1
<i>počet odpovědí s explicitním ohodnocením</i>	
<i>přehled všech odpovědí</i>	• ra 2 • ra 3 • ra 2 • ra 2 • ra 4 • ra 4 • ra 1 • ra 2 • ra 2 • ra 2 • ra 3 • ra 3 • ra 2 • ra 4 • ra 2 • ra 2 • ra 2 •
<i>počty různých odpovědí</i>	1x ra 1 1x ra 2 10x ra 3 3x ra 4 3x
<i>qdef soubor</i>	/el/1421/jaro2014/PSA_038/odp/tb/test03.qdef

Psychometrická analýza položek

❏ Diskriminační účinnost

- ❏ je nejvyšší, pokud položku řeší správně osoby, které dosahují vysokého celkového skóru a pokud ji nedokážou vyřešit osoby, které dosahují nízkého celkového skóru
- ❏ lze ji spočítat jako (korigovanou) korelaci skóru položky a celkového skóru (v SPSS v rámci výpočtu vnitřní konzistence)

❏ Nízká rozlišovací účinnost ($r=-0,03$)

- ❏ Dětský mléčný chrup je tvořen dvaceti zuby. Oproti chrupu dospělého člověka chybí
 - ❏ a) **zuby třenové** b) řezáky c) stoličky d) špičáky

❏ Vynikající rozlišovací účinnost ($r=0,77$)

- ❏ Podíl geneticky podmíněné variance a celkové fenotypové variance znaku se označuje
 - ❏ a) **koeficient dědivosti** b) expresivita alel c) koeficient hybridizace d) intermediarita

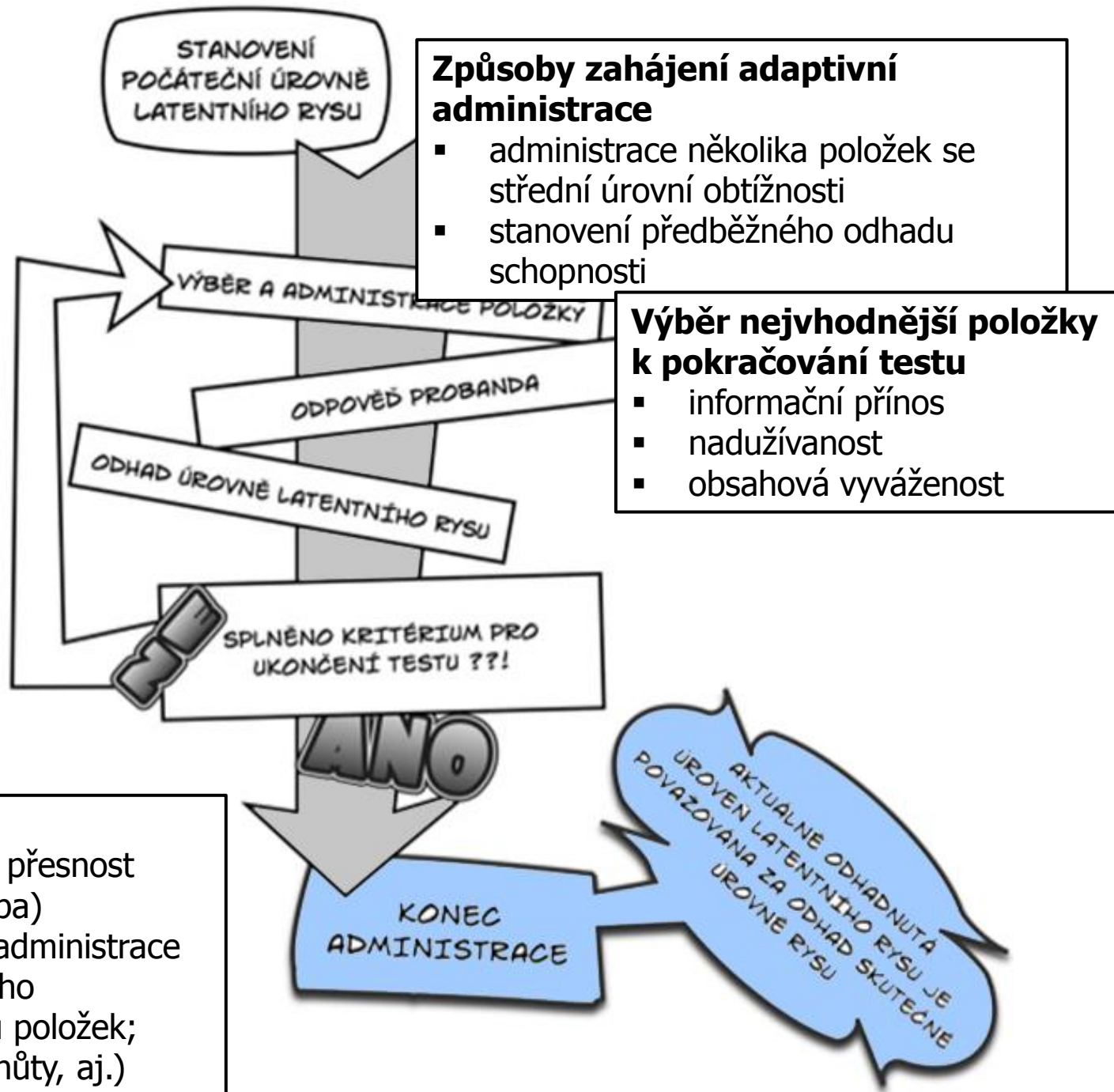
Rozpoznání nevhodných úloh

- nízká (záporná) diskriminační schopnost
 - příliš nízká obtížnost u testů s účelem rozlišovacím
 - příliš vysoká obtížnost a současně nízká diskriminační schopnost
 - vysoký podíl neodpovězeno a současně nízká diskriminační schopnost
 - nefunkční distraktory
- není vždy nutné položku vyřadit z testové banky, někdy stačí modifikace zadání, distraktorů...

Adaptivně administrované testy

- ❏ **výběr testovacích úloh se přizpůsobuje schopnostem testované osoby**
- ❏ základem testu je kvalitní položková banka
 - ❏ soubor položek, které se vztahují k určitému konstrukt, a jsou známy jejich parametry (např. obtížnost)
 - ❏ obsahuje velké množství položek
 - ❏ dostatečně pokrývá celé spektrum obtížnosti
 - ❏ obsahuje dostatečné množství úloh v každé tematické oblasti testovaného konstrukt
- ❏ v každém kroku adaptivní administrace je přepočítán (upřesněn) odhad schopnosti testované osoby; na základě tohoto odhadu a známých parametrů položek je vybrána nejvhodnější položka

CAT



Závěr testu

- dostatečně vysoká přesnost odhadu (nízká chyba)
- speciální případy (administrace předem stanoveného maximálního počtu položek; vyčerpání časové lhůty, aj.)

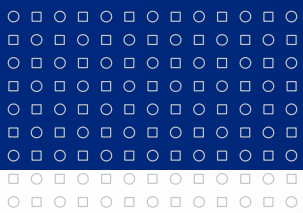
Adaptivní administrace testu

➤ výhody

- efektivita
- možnost opakované administrace testu bez efektu učení
- flexibilita
- udržení motivace administrací vhodných položek

➤ nevýhody

- nároky na technické vybavení (zejména vývoj software či jeho pořízení)



Děkujeme za pozornost.

