

Soubor postupů a nástrojů pro zkvalitnění tvorby znalostních testů pomocí psychometrických modelů

Patricia Martinková, Eva Potužníková, Jan Netík a kol.



PEDAGOGICKÁ FAKULTA
Ústav výzkumu a rozvoje vzdělávání
Univerzita Karlova



Ústav informatiky
Akademie věd ČR



Projekt TL05000008 Výzvy pro hodnocení znalostí: Analytická podpora tvorby znalostních testů
byl spolufinancován se státní podporou Technologické agentury ČR v rámci Programu ÉTA 5.

Obsah

Psychometrická analýza maturitních a jiných testů pomocí interaktivní aplikace ShinyItemAnalysis a modulu „EduTest Item Analysis“ (popis metod a návod na implementaci) <i>Patricia Martinková, Jan Netík, Adéla Hladká</i>	2
Predikce obtížnosti položek pomocí modulu „EduTest Text Analysis“ (popis metod a návod na implementaci) <i>Jana Dlouhá, Jan Netík, Lubomír Štěpánek, Eva Potužníková, Patricia Martinková</i>	31
Analýza školních dat pomocí Interaktivní aplikace pro střední školy „EduTest maturita“ (popis metod a návod na implementaci) <i>Jan Netík, Eva Potužníková, Patricia Martinková</i>	54
Počítačové adaptivní testování v rámci aplikace „EduTest CAT“ (popis metod a návod na implementaci) <i>Iván Leonardo Pérez Cabrera, Jan Netík, Eva Potužníková, Patricia Martinková</i>	77
EduTest Item Analysis: Modul pro analýzu položek znalostních testů (software) <i>Jan Netík, Patricia Martinková</i>	89
EduTest Text Analysis: Modul pro predikci obtížnosti položek znalostních testů z jejich textového zadání (software) <i>Jan Netík, Jana Dlouhá, Patricia Martinková, Lubomír Štěpánek</i>	90
EduTest Maturita: Interaktivní aplikace pro analýzu dat z didaktických testů maturitní zkoušky (software) <i>Jan Netík, Eva Potužníková, Patricia Martinková</i>	91
EduTest JPZ: Interaktivní aplikace pro analýzu dat z didaktických testů jednotné přijímací zkoušky (software) <i>Jan Netík, Eva Potužníková, Patricia Martinková</i>	92
EduTestCAT: Interaktivní aplikace k procvičování maturitních úloh s podporou počítačového adaptivního testování (software) <i>Iván Leonardo Pérez Cabrera, Jan Netík, Patricia Martinková</i>	93

Počítačové adaptivní testování v rámci aplikace EduTest CAT

Iván Leonardo Pérez Cabrera, Jan Netík,
Eva Potužníková, Patrícia Martinková

1 Úvod

Počítačové adaptivní testování (*computerized adaptive testing* – CAT) je způsob zadávání testů, v němž je výběr testových položek přizpůsoben úrovni měřeného rysu každého respondenta. Hlavním účelem CAT je nabídnout efektivnější a přesnější měření, než jakého by bylo dosaženo pomocí lineárního testu, v němž jsou všem respondentům zadávány stejné položky, obvykle i ve stejném pořadí (van der Linden & Glas, 2010). CAT se stalo důležitou oblastí výzkumu a je úspěšně využíváno například v oblasti psychologie, psychiatrie a medicíny pro měření psychických rysů, schopností či míry subjektivních zdravotních potíží a prosazuje se i v pedagogickém testování při měření znalostí. Vzhledem k zaměření tohoto dokumentu na oblast pedagogického testování budeme nadále hovořit o měření znalostí, principy CAT lze však analogicky aplikovat i na jiné oblasti měření.

V rámci CAT jsou testovaným osobám zadávány položky v pořadí, které se přizpůsobuje výkonu daného jednotlivce, aby se dosáhlo přesnějšího odhadu měřených znalostí. Pořadí položek je optimalizováno tak, aby každá vybraná položka přinášela v aktuálním kroku testu co nejvíce informací o úrovni měřených znalostí.

Oproti tradičnímu (lineárnímu) testování má CAT několik výhod. Zejména vede ke zkrácení testů pro testované osoby, neboť je nezatěžuje nadbytečným množstvím položek, které by pro ně byly příliš snadné nebo příliš obtížné. I s menším počtem položek měří jejich znalosti s dostatečnou přesností (Štěpánek & Martinková, 2020), protože položky jsou vybírány tak, aby byly při měření daného respondenta co nejefektivnější. To je podstatné zvláště pro respondenty s extrémně nízkou či extrémně vysokou úrovní znalostí, které klasické lineární testy navržené pro běžnou populaci měří méně přesně než průměrné respondenty. CAT také zpravidla zvyšuje motivaci během testování, protože předkládané položky odpovídají individuální úrovni testovaných a jsou pro ně relevantní. Snižuje také riziko podvádění, protože každý respondent má jiný výběr a jiné pořadí položek (Martinková & Hladká, 2023).

V roce 2023 vyvinulo Centrum pro zjišťování výsledků vzdělávání (CZVV) novou webovou aplikaci *Trénuj a uč se* (TAU)¹ k procvičování přijímacích testů, která má být do budoucna rozšířena také o procvičování maturitních testů. Aplikace obsahuje otázky použité u zkoušek z předchozích let. Její užívání je zdarma a uživatel se nemusí přihlašovat. K procvičování si uživatel může vybrat celý test nebo skupinu položek a jeho výsledky jsou vyhodnoceny online. Jakmile jsou vybrány položky k procvičování, jejich zadávání je sekvenční (lineární).

Ve stejné době vytvořil realizační tým projektu EduTest aplikaci nazvanou *EduTest CAT* se stejným cílem: poskytnout uživatelům nástroj k procvičování testů pomocí sad otázek z předchozích let. Na rozdíl od aplikace TAU je *EduTest CAT* adaptivním testováním, které výběr zadávané položky přizpůsobuje aktuálnímu výkonu uživatele vyhodnocovanému pomocí moderních psychometrických modelů. V současné době aplikace obsahuje položky z maturitních testů z matematiky zadávané v letech 2019 a 2022, do budoucna může být rozšířena o další roky a předměty, případně také o testy z jednotných přijímacích zkoušek.

V následující sekci 2 stručně popíšeme principy počítačového adaptivního testování, v sekci 3 pak podrobně vysvětlíme, jak jsou tyto principy implementovány v aplikaci *EduTest CAT*.

¹<https://tau.ceremat.cz/>

2 Principy adaptivního testování

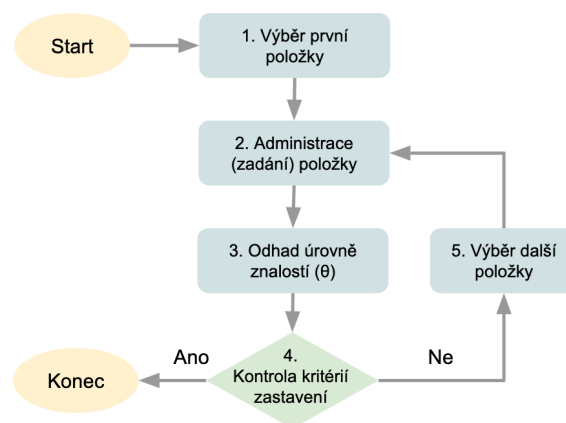
Základem CAT je banka kalibrovaných položek (tj. položek se známými psychometrickými vlastnostmi), z níž se vytváří konkrétní test pro každého respondenta. Dalšími strukturálními prvky CAT jsou proces výběru vhodných položek v každém kroku testování, proces odhadu úrovně znalostí respondenta po zadání každé položky a kritéria zastavení.

V rámci CAT se pro odhady psychometrických vlastností položek i pro odhady znalostí testovaných osob používají modely teorie odpovědi na položku (*item response theory* – IRT). Pro IRT modely použité v naší CAT aplikaci platí dva základní předpoklady: 1) měřená znalost (označovaná řeckým písmenem θ) je jednorozměrná; jinými slovy, test měří pouze jeden rys a 2) odpovědi na položky jsou navzájem lokálně nezávislé; jinými slovy, (ne)úspěšné řešení jedné položky nemá vliv na výsledek řešení jiné položky (Magis et al., 2017).

2.1 Průběh adaptivního testování

Proces adaptivního testování lze schematicky rozložit do pěti po sobě jdoucích kroků znázorněných na obr. 1 a popsaných níže.

Obrázek 1: Vývojový diagram CAT



1. Výběr první položky. První krok spočívá ve výběru jedné položky pro zahájení adaptivního testu. Na začátku testování zpravidla nemáme informace o znalostech testované osoby, takže jako první položka je obvykle vybrána ta, která poskytuje nejvíce informací o znalostech průměrného respondenta. Při výběru první položky je však možné zohledňovat i jiné faktory, např. podíl správných odpovědí v předchozích testováních pro zvýšení motivace na začátku testu nebo předpokládanou úroveň znalostí na základě dodatečných informací o respondentovi.

2. Administrace (zadání) položky. V tomto kroku se položka zobrazí uživateli a vyhodnotí se jeho odpověď.

3. Odhad úrovně znalostí. V tomto kroku se po každé administraci položky vypočítá odhad úrovně znalostí θ daného respondenta, přičemž se berou v úvahu jeho odpovědi na všechny dosud řešené položky.

4. Kontrola kritérií zastavení. V tomto kroku se ověřuje, zda byla dosažena kritéria zastavení. Pokud ne, je vybrána a zadána nová položka. Pokud ano, proces adaptivního

testování skončí. Kritéria zastavení mohou být definována například s ohledem na požadovanou přesnost měření, maximální povolenou délku testu či jiné faktory (viz část 3.2.4).

5. **Výběr další položky.** Pokud nejsou splněna kritéria zastavení, vybere se další položka, která je v daném kroku vyhodnocena jako optimální. Existuje mnoho způsobů, jak vybrat další položku. Aplikace *EduTest CAT* používá kritérium maximální Fisherovy informace (MFI), které spočívá ve výběru té položky, která pro aktuální odhad úrovně znalostí θ podává maximální množství informace.

Smyčka kroků 2–5 se opakuje, dokud nejsou splněna kritéria pro zastavení. Poté již nejsou nabízeny žádné další položky a aktuální odhad úrovně znalostí se přijme jako výsledný.

3 Online aplikace *EduTest CAT*

Aplikace *EduTest CAT* využívá k implementaci funkcí adaptivního testování softwarovou knihovnu *catR* (Magis & Barrada, 2017) napsanou v jazyce R.

Položkovou banku aplikace *EduTest CAT* tvoří sady úloh z maturitních testů z matematiky zadávaných v jarních termínech 2019 a 2022. Pro jejich kalibraci jsme použili dvouparametrický logistický (2PL) IRT model, který na základě anonymizovaných položkových dat dostupných na webu CZVV² určil pro každou položku parametry diskriminace (parametr a) a obtížnosti (parametr b). Tyto dva parametry jsou klíčové pro výběr nejvhodnější položky v každém kroku adaptivního testování. Parametr obtížnosti určuje, pro jakou úroveň znalostí přináší položka nejvíce informace, zatímco parametr diskriminace vyjadřuje množství informace poskytované položkou.

Aplikace *EduTest CAT* je dostupná na následující adrese:

<https://edutest.cs.cas.cz/cat/>

Na první stránce se zobrazí ověřovací formulář (obrázek 2), v němž je nutné zadat uživatelské jméno a heslo (např. uživatelské jméno: **catlogin1**, heslo: **catpassword1**).

Obrázek 2: Přihlašovací stránka do aplikace



²<https://data.ceremat.cz/menu/maturitni-zkouska>

3.1 Domovská stránka

Po přihlášení aplikace zobrazí úvodní domovskou stránku, kde si uživatel vybere testy a témata, která ho zajímají (obrázek 3). Uživatel si musí vybrat alespoň jeden test, aby měl k dispozici sadu otázek; jinak se zobrazí výstražná zpráva.

Obrázek 3: Výběr testů a témat

Testy	<input type="checkbox"/>
Maturitní zkouška 2019-jaro	<input type="checkbox"/>
Maturitní zkouška 2022-jaro	<input checked="" type="checkbox"/>
Témata	<input type="checkbox"/>
Číselné množiny	<input checked="" type="checkbox"/>
Algebraické výrazy	<input checked="" type="checkbox"/>
Rovnice a nerovnice	<input checked="" type="checkbox"/>
Funkce	<input checked="" type="checkbox"/>
Posloupnosti a řady, finanční matematika	<input type="checkbox"/>
Planimetrie	<input type="checkbox"/>
Stereometrie	<input type="checkbox"/>
Analytická geometrie	<input type="checkbox"/>
Kombinatorika, pravděpodobnost, statistika	<input type="checkbox"/>

V současné době obsahuje webová aplikace dva testy odpovídající maturitním zkouškám z matematiky pro jarní zkušební termíny 2019 a 2022. Každý test obsahuje 30 položek, které byly rozděleny do 9 témat podle tematických okruhů dostupných na webu CZVV.³ Položky ze zvolených testů a témat tvoří aktuální banku položek pro CAT.

Je důležité zmínit, že v aplikaci *EduTest CAT* jsou odpovědi na všechny položky hodnoceny binárně (1 = zcela správně, 0 = nesprávně nebo částečně správně), na rozdíl od původního hodnocení maturitní zkoušky, v níž lze za položku získat až 4 body v závislosti na její složitosti. Důvodem je to, že dvouparametrický IRT model použitý pro kalibraci položek a odhad úrovně znalostí respondentů předpokládá binární data.

3.2 Implementace kroků CAT v aplikaci *EduTest CAT*

Jakmile si uživatel vybere kombinaci testů a témat a svůj výběr potvrdí kliknutím na tlačítko **Odeslat**, začne adaptivní testování, jehož průběh je popsán níže.

³https://maturita.cermat.cz/files/files/katalog-pozadavku/MX-katalog_pozadavku.pdf

3.2.1 Výběr první položky

Pohled na uživatelské rozhraní aplikace během adaptivního testování je znázorněn na obrázku 4. Hlavní rozhraní obsahuje tři panely: 1) znění aktuální položky, 2) sekvenci postupných odhadů znalosti respondenta a 3) informační křivky položek (*item information curves*, IIC).

V panelu 1 na obrázku 4 vidíme znění první zadané položky. V panelu 2 na témže obrázku vidíme úvodní odhad znalosti respondenta $\hat{\theta} = 0$. V panelu 3 na obrázku 4 jsou jednotlivé položky z aktuální banky položek reprezentovány křivkami v grafu, jehož svislá osa odpovídá množství informace poskytované položkou a vodorovná osa odpovídá úrovni měřené znalosti θ . Informační křivka položky ukazuje, kolik informace o úrovni znalostí položka poskytuje pro každou úroveň θ .

Informační křivky položek mají zvonovitý tvar, který je dán hodnotami parametrů IRT modelu použitého k jejich kalibraci. Vrchol informační křivky se nachází v hodnotě, která odpovídá parametru obtížnosti položky. S rostoucí vzdáleností od této hodnoty položka poskytuje čím dál méně informací. Výšku a strmost křivky určuje parametr diskriminace položky. Položky, které mají velkou diskriminaci, mají křivky rychle vystoupavé a úzké. Takové položky mají velkou informační hodnotu při zjišťování znalostí respondenta v relativně úzké oblasti jejich obtížnosti. Naopak mimo tuto oblast mají velmi malou informační hodnotu, jinými slovy nepomáhají při měření znalostí respondentů jejichž úroveň znalostí leží výrazně mimo vrchol křivky. Položky s malou diskriminací (široké křivky bez výrazného vrcholu) jsou schopny měřit znalosti respondentů v širší oblasti úrovně znalostí, byť s menší přesností

Jako první položka je vybrána ta, která přináší nejvíce informací pro průměrnou úroveň znalostí předchozí populace (průměrná úroveň znalostí se v rámci IRT modelování běžně nastavuje na hodnotu nula). V panelu 3 na obr. 4 je vidět, že pro průměrnou hodnotu znalostí předchozí populace (hodnota 0 na vodorovné ose) poskytuje nejvíce informací položka 10 v bance (červená informační křivka). Informační křivka této položky protíná svislou linku vedenou z bodu 0 na vodorovné ose výše než informační křivky ostatních položek. Tato položka je tedy vybrána jako první.

3.2.2 Administrace položky

Položky jsou uživateli prezentovány v rozhraní aplikace (panel 1 na obr. 4). V tomto kroku se položka zobrazí a odpověď uživatele se vyhodnotí podle typu položky.

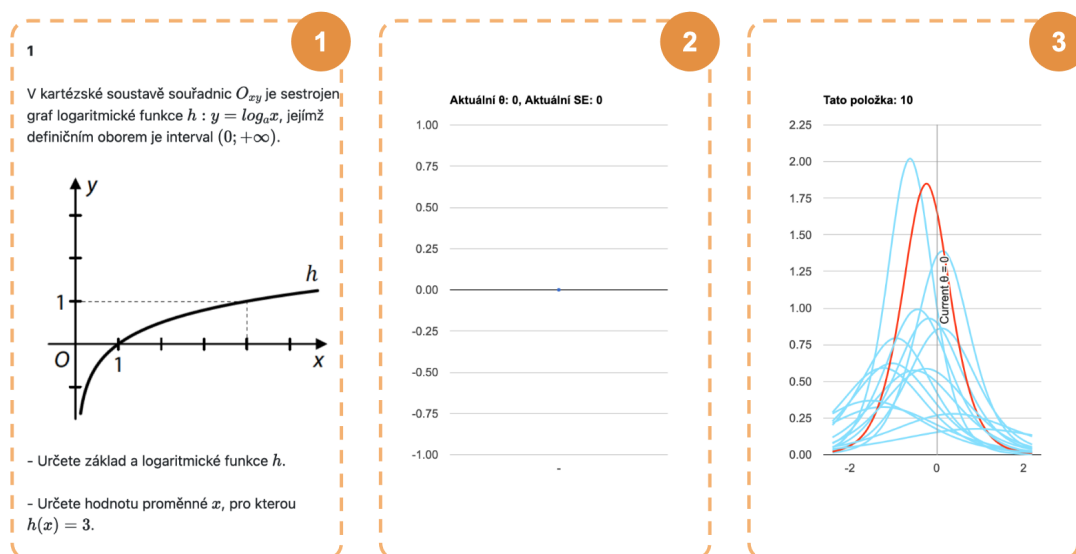
V bance existují dva typy položek: uzavřené otázky, v nichž uživatel vybírá správnou odpověď z několika nabízených možností, a otevřené otázky, v nichž uživatel svou odpověď samostatně tvoří. Při ostrém maturitním testu používá testovaný žák k zodpovězení těchto otázek tužku a papír a jeho odpověď je hodnocena proškolenými hodnotiteli na základě jednotných kritérií.

V aplikaci *EduTest CAT* je odpověď na uzavřené otázky vyhodnocena automaticky poté, co uživatel vybere jednu možnost a svůj výběr potvrdí kliknutím na tlačítko **Odeslat** (obrázek 5). Pokud je položka otevřená, musí uživatel svoji odpověď vyhodnotit sám porovnáním se správným řešením zobrazeným v rozhraní (obrázek 6).

3.2.3 Odhad úrovně znalostí

Jakmile uživatel na položku odpoví a jeho odpověď je vyhodnocena, systém odhadne aktuální úroveň znalostí uživatele na základě jeho odpovědí na všechny dosud adminis-

Obrázek 4: Hlavní rozhraní procesu CAT: 1) aktuální položka, 2) sekvence postupných odhadů znalostí, 3) informační křivky položek



10

V geometrické posloupnosti je třetí člen $a_3 = 2$

a čtvrtý člen je o 3 menší než třetí člen.

Jaký je součet prvních tří členů uvedené geometrické posloupnosti $(a_1 + a_2 + a_3)$?

- 6
- jiný součet
- 3
- 26
- 15

Odeslat

Obrázek 5: Příklad uzavřené položky

trované položky.

V panelu 2 na obrázku 4 je graficky zobrazena aktuální trajektorie odhadů znalostí (detail viz obrázek 7). Svislá osa představuje úroveň znalostí θ , zatímco vodorovná osa

9

Je dán interval A a množina B :

$$A = \langle -5; 5 \rangle, B = \{x \in \mathbb{R}; -8 \leq x < 3\}$$

Určete $A \cap B$.

(a)

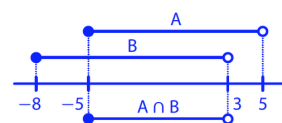
Řešení:

✕

$$A = \langle -5; 5 \rangle$$

$$B = \langle -8; 3 \rangle$$

$$A \cap B = \langle -5; 3 \rangle$$



Vaše řešení bylo správné?

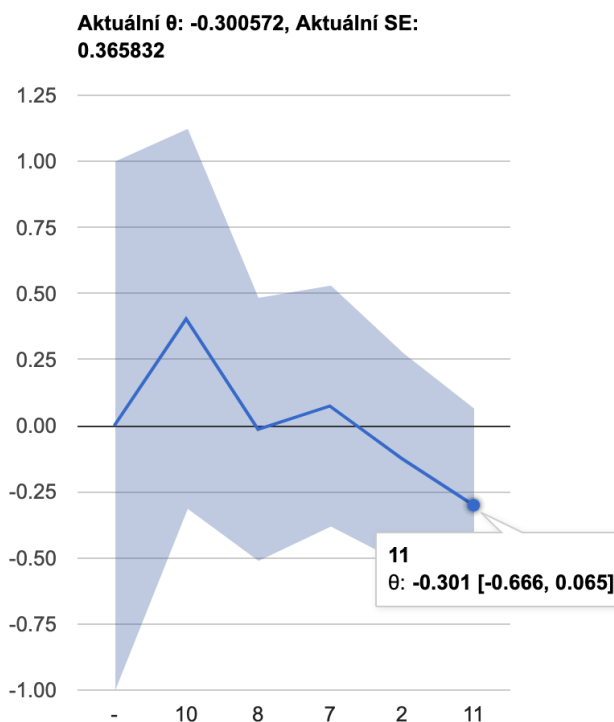
(b)

Obrázek 6: Příklad otevřené položky: (a) zadání, (b) řešení a sebehodnocení

označuje položky, které již byly zadány a vyhodnoceny. Modrou čarou je znázorněna sekvence odhadů znalostí v jednotlivých krocích CAT, přičemž šířka šedého pruhu kolem odhadů znalostí vyjadřuje interval spolehlivosti těchto odhadů.

Příklad znázorněný na obr. 7 ilustruje, že odhadovaná úroveň znalostí během adaptivního testování roste, když uživatel odpoví na zadané položky správně (otázky 10 a 7), a klesá, když na zadané otázky odpoví špatně (otázky 8, 2, 11). Zároveň je vidět, že rozpětí intervalů spolehlivosti s délkou testu klesá (v důsledku souvisejícího snižování standardních chyb (SE) odhadu znalostí). Aktuální odhad úrovně znalostí včetně intervalu spolehlivosti po každé administraci položky je uveden také v rozbalovacím políčku po najetí myší (obr. 7).

Obrázek 7: Trajektorie odhadů znalostí



3.2.4 Kontrola kritérií zastavení

V tomto kroku se ověřuje splnění kritérií zastavení a dojde k rozhodnutí, zda bude vybrána a zadána nová položka, nebo zda bude adaptivní testování ukončeno. V CAT se zpravidla používá jedno ze čtyř kritérií zastavení (van der Linden & Glas, 2010), nebo jejich kombinace:

- a) kritérium délky – je stanoven maximální celkový počet položek, které mají být administrovány, a proces CAT se zastaví po dosažení tohoto počtu,
- b) kritérium času – je stanoven celkový čas, a proces CAT se zastaví po uplynutí tohoto času,
- c) kritérium přesnosti – CAT se zastaví, když chyba měření klesne pod předem stanovenou mez,
- d) kritérium klasifikace – je stanovena mez pro klasifikaci, např. minimální požadovaná úroveň zvládnutí testu, a CAT se zastaví, když interval spolehlivosti odhadu znalosti je mimo tuto mez, a je tedy s dostatečnou přesností rozhodnuto o jejím splnění či nesplnění.

V aktuální verzi online aplikace *EduTest CAT* je použita kombinace dvou kritérií zastavení:

1. standardní chyba odhadu znalosti je menší než 0,3;
2. maximální počet položek předložených uživateli je 10.

Test se zastaví, pokud bude splněno alespoň jedno z těchto kritérií. Současně platí, že CAT je ukončeno, pokud byly vyčerpány všechny dostupné položky, tedy pokud je v počátečním výběru testů a témat počet položek nižší než 10, může být test zastaven ještě před splněním jednoho z výše uvedených kritérií.

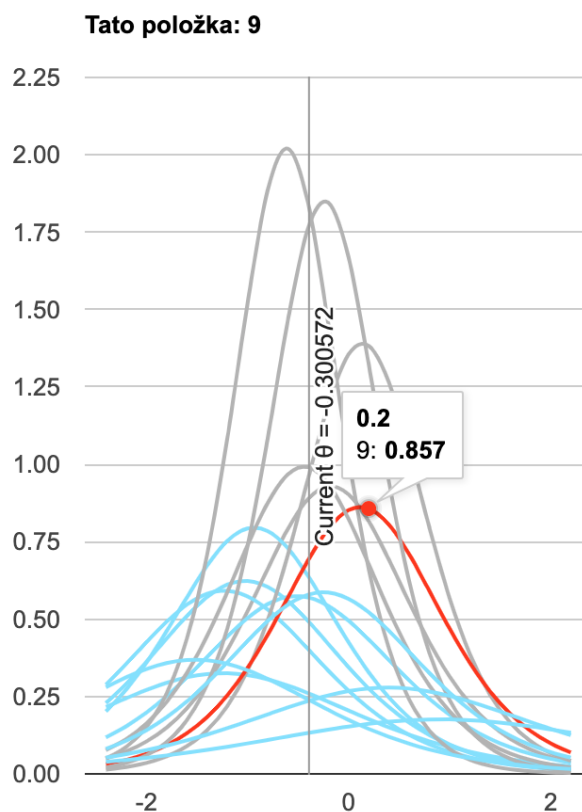
3.2.5 Výběr další položky

Další položka se vybere z aktuální banky položek s ohledem na aktuální odhad znalostí respondenta, který je založen na jeho předchozích odpovědích (*response pattern*). Konkrétně se vybere ta položka, která pro hodnotu aktuálního odhadu znalostí respondenta poskytuje nejvíce informací. Položky, které již byly administrovány v předchozích krocích CAT, nemohou být znovu použity, optimální položka se tedy vybírá pouze ze zbývajících položek ze zvolených testů a témat.

Při výběru položek se vychází z jejich informačních křivek (IIC), které jsou zobrazeny v rozhraní aplikace (panel 3 na obr. 4). Červená křivka odpovídá aktuální položce (obrázek 8).

V příkladu znázorněném na obr. 8 je aktuální odhad úrovně znalostí uživatele $-0,3$ (svislá linka v grafu). Byla mu proto vybrána položka 9 (červená křivka) jako nejvíce informativní, neboť průsečík její IIC se svislou linkou vycházející z bodu $-0,3$ na vodorovné ose leží ze všech dostupných položek nejvýše. V bance položek jsou sice dostupné také položky, které mají vrchol svých IIC blíže k hodnotě $-0,3$, avšak tyto položky jsou méně informativní (mají nižší hodnoty diskriminačního parametru), a proto nejsou v tomto kroku vybrány.

Obrázek 8: Informační křivky položek (IIC)



3.3 Konec adaptivního testu

Po ukončení adaptivního testování se zobrazí závěrečné hodnocení, jak je znázorněno na obr. 9. V tomto hodnocení je vidět počet zadaných a počet správně zodpovězených položek, odhadovaná úroveň znalostí a také standardní chyba měření. Zobrazena je také kompletní trajektorie odhadů znalostí.

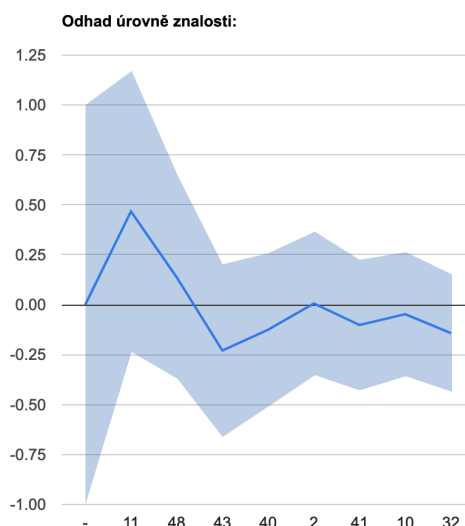
V prezentovaném příkladu bylo respondentovi předloženo 8 položek. V každém kroku se standardní chyba odhadu snižovala, až nakonec byla nižší než stanovená prahová hodnota 0,3, kdy adaptivní test skončil. Konečný odhad úrovně znalostí respondenta byl $-0,14$ (SE 0,29).

4 Závěr

V tomto dokumentu jsme popsali fungování webové aplikace *EduTest CAT*, nástroje pro procvičování na maturitní zkoušku. Procvičování v aplikaci *EduTest CAT* je adaptivní, jednotlivé položky jsou vybírány s ohledem na úspěšnost uživatele při řešení dříve zadaných položek. Ve srovnání s lineárně řešenými testy může být pro uživatele procvičování v adaptivní aplikaci více motivující, protože předkládané položky pro něj nejsou ani příliš jednoduché ani příliš složité, ale představují pro každého respondenta adekvátní výzvu. V rámci přípravy na maturitní zkoušku je nicméně vhodné aplikaci kombinovat také s řešením kompletních testů, aby se uživatel dostatečně seznámil se všemi typy otázek, které

Obrázek 9: Závěrečné hodnocení

Uživatel: **catlogin1**
 Počet zodpovězených položek: **8**
 Počet správných odpovědí: **4**
 Odhadnutá úroveň znalosti: **-0.141952**
 Střední chyba: **0.293675**



se u zkoušky mohou vyskytnout.

Kromě možnosti adaptivního procvičování maturitních otázek aplikace názorně demonstrovuje hlavní principy CAT. To rozšiřuje její možné využití také např. pro osvětlení podstaty adaptivního testování široké veřejnosti.

Aktuální verze aplikace vnikla jako pilotní produkt, který je možné dále rozvíjet v několika směrech:

V současné době se položková banka dostupná v aplikaci *EduTest CAT* skládá z 60 položek (30 z jarního termínu 2019 a 30 z jarního termínu 2022). Tuto banku lze rozšířit o položky z dalších zkušebních termínů. V případě položek z podzimních termínů je potřeba zajistit, aby odhadnuté hodnoty jejich parametrů nebyly vychýleny z důvodu výrazně nižších znalostí žáků, kteří skládají zkoušku na podzim.

Položková banka obsahuje dva typy položek: uzavřené a otevřené. Uzavřené položky jsou automaticky vyhodnocovány aplikací, zatímco otevřené položky musí vyhodnotit sám uživatel porovnáním svého řešení s řešením prezentovaným v rozhraní aplikace. Pro pilotní provoz to je dostačující, více automatizované hodnocení by však užívání aplikace podstatně zjednodušilo. V budoucnu by pro otevřené položky mohla být začleněna interakce hodnotitele, případně by mohlo být u některých úzce otevřených položek využito automatické hodnocení.

V aktuální verzi aplikace *EduTest CAT* je jako první položka vybrána ta, která je nejvíce informativní pro úroveň znalostí blízkou průměrné znalosti předchozí populace. To znamená, že pro danou kombinaci vybraných testů a témat se jako první zobrazí vždy stejná položka, některé položky jsou tedy vybírány častěji než jiné. V budoucnu lze aplikaci rozšířit o tzv. řízení míry expozice položek (Chang & Ansley, 2003), které zajistí,

aby žádná položka nebyla zobrazována příliš často. K tomu by však bylo potřeba zajistit, aby si aplikace pamatovala odpovědi předchozích uživatelů nebo daného přihlášeného uživatele. Při výběru první položky může pak být také například zohledňována úroveň znalostí respondenta naměřená v předchozích případech užití aplikace.

V současné době aplikace dovoluje před spuštěním testu vybrat z nabízených témat, ale zastoupení položek z vybraných témat není nijak řízeno. Další možností by bylo doplnit do aplikace tzv. vyvažování obsahu, kdy je kromě optimalizace informace užití položky navíc přibližně zajištěno požadované zastoupení jednotlivých témat. Respondent by si tak mohl stanovit, že požaduje cvičný test obsahující 30 % položek z geometrie apod.

Aplikace *EduTest CAT* využívá jednodimenzionální IRT model (měřící jeden rys, zde znalosti z matematiky), který je standardem v mnoha testových nástrojích založených na IRT. V budoucnu by bylo možné uvažovat také o vícerozměrném modelu (van Groen et al., 2019), který může vedle matematických znalostí zohlednit např. také verbální složku, která může hrát roli při řešení matematických slovních úloh.

Literatura

- Chalmers, R. P. (2016). Generating Adaptive and Non-Adaptive Test Interfaces for Multidimensional Item Response Theory Applications. *Journal of Statistical Software*, 71(5), 1–39. <https://doi.org/10.18637/jss.v071.i05>
- Chang, S.-W., & Ansley, T. N. (2003). A comparative study of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, 40(1), 71–103. <https://doi.org/10.1111/j.1745-3984.2003.tb01097.x>
- Magis, D., & Barrada, J. R. (2017). Computerized Adaptive Testing with R: Recent Updates of the Package catR. *Journal of Statistical Software, Code Snippets*, 76(1), 1–19. <https://doi.org/10.18637/jss.v076.c01>
- Magis, D., Yan, D., & Von Davier, A. A. (2017). *Computerized Adaptive and Multistage testing with R: Using Packages catR and mstR*. Springer. <https://doi.org/10.1007/978-3-319-69218-0>
- Martinková, P., & Hladká, A. (2023). *Computational Aspects of Psychometric Methods: With R (1st ed.)* Chapman; Hall/CRC. <https://doi.org/10.1201/9781003054313>
- Štěpánek, L., & Martinková, P. (2020). Feasibility of computerized adaptive testing evaluated by Monte-Carlo and post-hoc simulations. *Proceedings of the 2020 Federated Conference on Computer Science and Information Systems*, 359–367. <https://doi.org/10.15439/2020f197>
- van der Linden, W. J., & Glas, C. A. W. (2010). *Elements of Adaptive Testing*. Springer. <https://doi.org/10.1007/978-0-387-85461-8>
- van Groen, M., Eggen, T., & Veldkamp, B. (2019, July). Multidimensional Computerized Adaptive Testing for Classifying Examinees. In *Theoretical and Practical Advances in Computer-Based Educational Measurement* (pp. 271–289). Springer Nature. https://doi.org/10.1007/978-3-030-18480-3_14