

SPRACOVANIE VÝSLEDKOV TESTU NOVOU TEÓRIOU TESTOV

THE INTERPRETATION OF TEST RESULTS BY THE NEW TEST THEORY

ĽUBOMÍR RYBANSKÝ, MARTA VRÁBELOVÁ

ABSTRAKT. *Cieľom tohto článku je poukázať na teóriu odpovedí na položky IRT a použiť ju pri spracovaní výsledkov výstupného testu projektu KEGA 3/7001/09 v školskom roku 2009/2010.*

KLÚČOVÉ SLOVÁ: *IRT, SG model, náročnosť úlohy, schopnosti*

ABSTRACT. *The aim of this paper is to mention the Item Response Theory and to use this theory for the treatment of the KEGA 3/7001/09 project outcoming test results from the school year 2009/2010.*

KEY WORDS: *IRT, SG model, item extremity, ability*

CLASSIFICATION: *MESC B10*

1. Úvod

Tento článok nadväzuje na článok [6], v ktorom boli klasickou teóriou testov spracované výsledky výstupného testu pre 5. ročník základnej školy vypracovaného v rámci projektu KEGA 3/7001/09 s názvom *Zvyšovanie kľúčových matematických kompetencií – alternatívne učebné programy z matematiky pre základné školy v zmysle cieľov nového štátneho vzdelávacieho programu a v zmysle zvyšovania matematickej gramotnosti podľa dopadov PISA*. Pripomíname, že projekt pokračuje projektom KEGA 015 UKF - 4/2012 a cieľom týchto projektov je vypracovať a experimentálne overiť nové, doplnujúce učebné materiály vo forme problémových úloh zameraných na riešenie problémov súčasného, bežného života pre 5. – 9. ročník ZŠ. Výsledky vstupného testu projektu obsahuje článok [5], a výsledky výstupného testu pre 6. ročník sú spracované v článku [3].

Novou teóriou testov použitou v tomto článku, nazývanou teória odpovedí na položky (Item Response Theory - IRT), odhadujeme charakteristické a informačné krivky pre jednotlivé úlohy výstupného testu pre 5. ročník ZŠ, úroveň matematických kompetencií pre každého žiaka, porovnáme túto úroveň pre experimentálne a kontrolné školy, a tiež pre skupiny škôl pri rozklade vzhľadom na skupinu a vyučovací jazyk. Najprv však čitateľovi priblížime podstatu IRT a predpoklady jej použitia. Teória odpovedí na položky bola použitá napr. v článku [7] s využitím balíka `ltm` programu *R* uvedeného v [4], a používa sa tiež pri spracovaní výsledkov testov PISA. Podrobnejšie informácie o IRT možno získať v [1], [8], [9], [10].

2. Podstata a predpoklady použitia IRT

Klasická teória testov je zameraná na deskriptívne charakteristiky skóre testu, na výpočet reliability testu, posúdenie validity testu, korelácie medzi položkami. Teória

odpovedí na položky je moderná teória testov (v porovnaní s klasickou teóriou testov), ktorá je veľmi populárna, pretože pomocou nej možno zistiť vlastnosti položiek a celého testu, ktoré klasická teória neposkytuje. Má však silnejšie predpoklady použitia. Prvým predpokladom je existencia spoločného faktora (latentnej premennej, schopnosti - ability) vysvetľujúceho korelácie položiek. Tento predpoklad nikdy nie je presne splnený, ale malé porušenia tohto predpokladu nespôsobujú veľké rozdiely. Stačí, keď pre údaje existuje jeden dominantný prvý faktor. Druhým predpokladom je, že vzťah medzi pozorovanými odpoveďami a latentnou premennou má špecifickú formu. Krivka znázorňujúca tento vzťah sa nazýva charakteristická krivka položky označovaná ako ICC. Latentná charakteristika sa označuje θ a predpokladá sa, že má normované rozdelenie, teda strednú hodnotu 0 a smerodajnú odchýlku 1. Charakteristická krivka je najčastejšie grafom logistickej funkcie. Na odhad logistickej funkcie sa v prípade testov s dichotomistickými odpoveďami (1 a 0) používa trojparametrický logistický model (niekedy menej alebo viacparametrický) tvaru

$$P(v_i = 1 | \theta = t) = c_i + \frac{1 - c_i}{1 + \exp(-1,7a_i(t - b_i))}.$$

Je to model pravdepodobnosti, že osoba s úrovňou latentnej premennej t odpovie na položku i správne (vyberie odpoveď 1). Parameter b_i sa nazýva parameter náročnosti (obťažnosti) i -tej položky, a_i je diskriminačný parameter a c_i je parameter hádania. Pre dobrý odhad parametrov charakteristickej krivky sú potrebné údaje veľkého rozsahu. Najdôležitejším parametrom je parameter náročnosti b_i . V bode b_i má krivka inflexný bod. Je to bod na osi x , pričom, ak $c_i = 0$, osoba so schopnosťou $\theta = b_i$ si vyberie danú odpoveď s pravdepodobnosťou 0,5. Čím je parameter b_i väčší, tým je položka náročnejšia. Náročnosť položky a schopnosť – ability sú na tej istej škále. Parameter diskriminácie a je tým väčší, čím je krivka strmšia, teda čím položka lepšie rozlišuje. Položky s malým diskriminačným parametrom slabo rozlišujú medzi osobami s rôznou úrovňou schopností a mali by sa z testu vylúčiť. Parameter c_i sa používa vtedy, keď sú v teste položky s výberom odpovede. Je to pravdepodobnosť, že osoba so žiadnou schopnosťou vyberie danú odpoveď. Výhodou IRT modelu je invariancia parametrov, čo znamená, že parametre nezávisia od výberu vzorky z populácie. Namiesto reliability v klasickej teórii testov sa v IRT používa informácia. Každá položka testu prispieva určitou informáciou do informácie testu. Táto informácia položky závisí od parametrov a a b položky, informácia testu je rovná súčtu informácií položiek. Uvedený trojparametrický model sa používa na reprezentáciu kognitívnych testov. Okrem tohto modelu sa používajú aj ďalšie modely.

3. Aplikácia IRT na výstupný test pre 5. ročník ZŠ

Výstupný test pre 5. ročník ZŠ bol použitý ako výskumný nástroj projektu KEGA 3/7001/09. Metódou výskumu bol experiment, školy zapojené do výskumu boli náhodne rozdelené na experimentálne a kontrolné. Výskumnú vzorku tvorí 877 žiakov 5. ročníka základných škôl zo štyroch okresov Nitrianskeho kraja. Niektoré školy sú školy s vyučovacím jazykom maďarským. Test obsahoval 6 úloh (položiek) obsahujúcich niekoľko podúloh. Všetky úlohy boli otvorené a boli bodované. Za každú zo 6 položiek žiak mohol získať 0 – 5 bodov.

Najprv sme overili splnenie predpokladov použitia IRT. Vypočítali sme korelačnú maticu položiek testu a zistili sme, že všetky korelácie sú významné. Metódou faktorovej analýzy sme našli jeden dominantný faktor vysvetľujúci 49,9% celkovej variability. Tento faktor nazveme *matematické kompetencie*.

Na spracovanie uvedeného testu je vhodné použiť *Samejima's Graded Response (SGR) Model* ([2]), čo je polytomický IRT model, ktorý predpokladá, že kódy (skóre, body) odpovedí na položky sú usporiadané. V tomto modeli sa pravdepodobnosť výberu k -tej odpovede (zisku k bodov) v i -tej položke (za i -tu úlohu) počíta podľa vzorca

$$P(v_i = k | \theta = t) = \frac{1}{1 + \exp(-1,7a_i(t - b_{i,k}))} - \frac{1}{1 + \exp(-1,7a_i(t - b_{i,k+1}))},$$

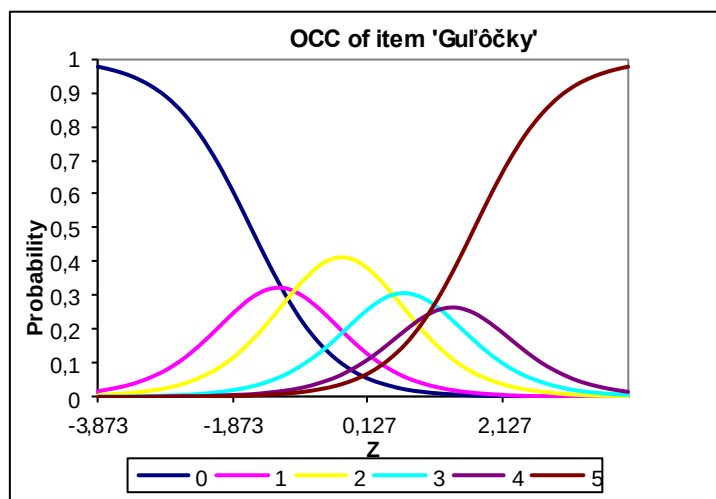
kde a_i je parameter diskriminácie, o ktorom sa predpokladá, že je rovnaký pre každý výber odpovede na i -tu položku, $b_{i,k}$ je parameter náročnosti k -tej odpovede na i -tu položku, pričom $b_{i,k-1} < b_{i,k} < b_{i,k+1}$, $k = 1, 2, \dots, s_i$, $b_{i,s_i+1} = \infty$, s_i je počet odpovedí na i -tu položku.

Na odhad parametrov modelu sme použili voľne šíriteľný program *eirt*, ktorý tvorí doplnok do Excelu. V tomto programe sme odhadli parametre charakteristických kriviek odpovedí na položky (získaných bodov) - OCC, uvedené sú v tabuľke 1. Najlepšie medzi žiakmi rozlišuje úloha *Kolko áut vyrobili* ($a_1 = 2,193$), nasledujú úlohy *Domy a ich čísla*, *Návšteva divadla* a *Gulôčky*, a najmenej rozlišujúcimi úlohami sú *Stavba* a *Požičovňa bicyklov*. Ak by sme náročnosť úloh posudzovali podľa parametra $b_{i,5}$, tak najnáročnejšími úlohami sú úlohy *Gulôčky* ($b_{4,5} = 1,720$) a *Požičovňa bicyklov* ($b_{2,5} = 1,565$). Menej náročnými sú úlohy *Kolko áut vyrobili* ($b_{1,5} = 0,485$) a *Domy a ich čísla* ($b_{5,5} = 0,435$). Najmenej náročnými úlohami sú *Stavba* ($b_{3,5} = 0,223$) a *Návšteva divadla* ($b_{6,5} = -0,067$). Nakreslili sme tiež grafy OCC pre jednotlivé počty bodov získané za úlohy. Úroveň matematických kompetencií je znázornená na tej istej škále (v intervale $[-4, 4]$) ako náročnosť. Ako ukážku uvádzame grafy OCC pre úlohu *Gulôčky* (obrázok 1) a *Návšteva divadla* (obrázok 2). V prípade úlohy *Gulôčky* (obrázok 1) žiak s úrovňou matematických kompetencií pod $-1,079$ najpravdepodobnejšie získa 0 bodov, žiak s úrovňou matematických kompetencií nad $-1,079$ a pod $-0,825$ najpravdepodobnejšie získa 1 bod, žiak s úrovňou matematických kompetencií nad $-0,825$ a pod $0,444$ najpravdepodobnejšie získa 2 body, žiak s úrovňou matematických kompetencií nad $1,206$ najpravdepodobnejšie získa 5 bodov. OCC pre ostatné úlohy, okrem *Návšteva divadla*, vyzerajú podobne. Z charakteristických kriviek úlohy *Návšteva divadla* (obrázok 2) je vidieť, že na celej škále latentnej premennej je najpravdepodobnejšie získať 0 bodov – úlohu nevyriešiť resp. 5 bodov – úlohu vyriešiť. Ostatné počty bodov majú relatívne malé pravdepodobnosti, čo naznačuje, že uvedenú úlohu by bolo vhodné prekódovať na dichotomickú premennú (0 - nesprávna odpoveď, 1 - správna odpoveď).

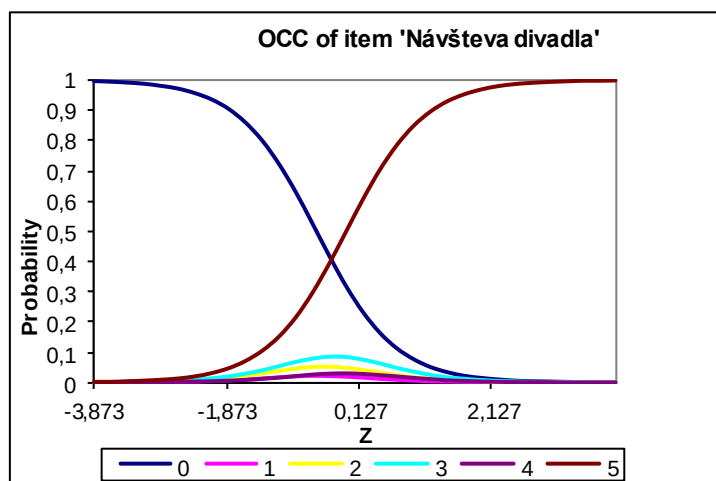
Testová informačná krivka je zobrazená na obrázku 3, celková informácia testu je rovná 21,26, informácia v intervale $[-4, 4]$ je rovná 20,92 (98,41%). Testové informačné krivky indikujú, že úlohy *Kolko áut vyrobili*, *Domy a ich čísla* a *Gulôčky* poskytujú najviac informácie (postupne 4,99, 4,53 a 4,44). Menej informácie sa získa z úloh *Požičovňa bicyklov* (2,73), *Stavba* (2,45) a *Návšteva divadla* (2,12). Výpočet množstva informácie sme urobili v programe *R*. Z predchádzajúceho vyplýva, že najmenej vhodnou úlohou testu je úloha *Návšteva divadla*.

	a_i	$b_{i,k}$					
		0	1	2	3	4	5
Koľko áut vyrobili...	2,193	-1,354	-1,006	-0,384	0,011	0,309	0,485
Požičovňa bicyklov	1,230	-1,545	-1,200	-0,355	0,699	1,409	1,565
Stavba	1,278	-1,964	-1,709	-1,059	-0,434	0,009	0,223
Gul'ôčky	1,642	-1,599	-1,191	-0,249	0,673	1,390	1,720
Domy a ich čísla	1,880	-1,929	-1,576	-1,053	-0,493	0,166	0,435
Návšteva divadla	1,695	-0,515	-0,490	-0,402	-0,239	-0,102	-0,067

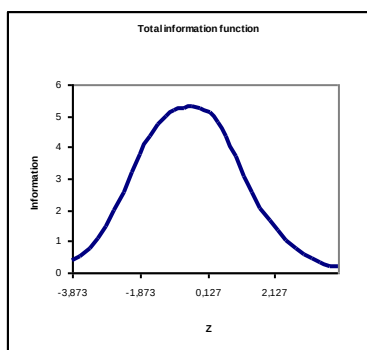
Tabuľka 1: Odhad parametrov charakteristických kriviek odpovedí



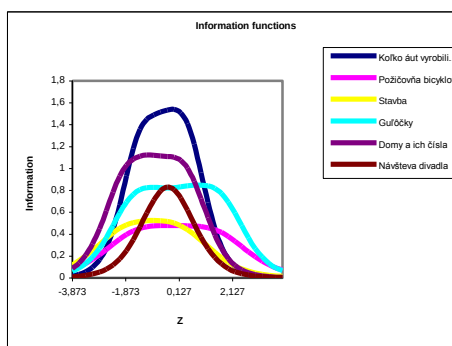
Obrázok 1: OCC pre *Gul'ôčky*



Obrázok 2: OCC pre *Návšteva divadla*



Obrázok 3: Testová informačná funkcia



Obrázok 4: Položkové informačné funkcie

Dôležitým výsledkom IRT je odhad matematických kompetencií žiakov. Pre vybraných šesť žiakov je odhad matematických kompetencií uvedený v tabuľke 2. Tu si môžeme všimnúť, že žiaci s rovnakou úrovňou matematických kompetencií (0,144) majú rozdielne skóre testu (18, 20) a traja žiaci, ktorí majú rovnaké skóre testu majú rozdielnu úroveň matematických kompetencií. Žiak číslo 552 s úrovňou matematických kompetencií 0,554 získal 20 bodov, pričom nevyriešil len najľahšiu úlohu *Návšteva divadla* a žiak číslo 457 s úrovňou matematických kompetencií 0,021 získal tiež 20 bodov, pričom za najťažšiu úlohu získal len jeden bod. Teda aj žiaci s veľmi rôznou úrovňou matematických kompetencií môžu získať rovnaký počet bodov.

žiak	skupina	VJ	Ú1	Ú2	Ú3	Ú4	Ú5	Ú6	Skóre	mat.kompetencie
516	E	SJ	2	2	1	3	5	5	18	0,144
633	E	SJ	1	2	5	2	5	5	20	0,144
112	K	SJ	5	5	5	3	5	5	28	1,307
423	E	MJ	5	5	2	5	5	5	27	1,409
552	E	SJ	5	2	5	3	5	0	20	0,554
457	E	MJ	4	3	5	1	2	5	20	0,021

Tabuľka 2

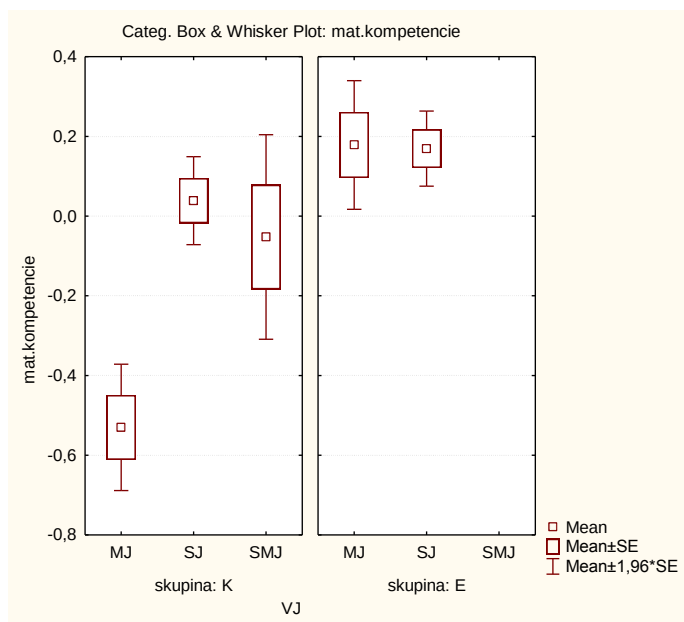
Vypočítané matematické kompetencie žiakov významne korelujú s bodovým hodnotením úloh. Spearmanove koeficienty korelácie obsahuje tabuľka 3.

Variable	Spearman Rank Order Correlations (mat.kompetencie) MD pairwise deleted Marked correlations are significant at $p < 0,05000$					
	Koľko áut vyrobili ...	Požičovňa bicyklov	Stavba	Gulôčky	Domy a ich čísla	Návšteva divadla
matematické kompetencie	0,809	0,596	0,596	0,736	0,756	0,672

Tabuľka 3: Korelácie matematických kompetencií žiakov a bodového hodnotenia úloh.

Aritmetický priemer matematických kompetencií je pre kontrolnú skupinu rovný -0,166, pre experimentálnu skupinu je rovný 0,172, smerodajné odchýlky sú 0,913 a 0,852. Rozdiel priemerov je významný (p -hodnota = 0,00000). Pri rozklade škôl vzhľadom na skupinu a vyučovací jazyk zistíme, že najvyššie priemerné kompetencie (0,179) dosiahli žiaci experimentálnych škôl s vyučovacím jazykom maďarským, nasledujú experimentálne školy s vyučovacím jazykom slovenským (0,169), kontrolné školy s vyučovacím jazykom

slovenským (0,038), kontrolné školy slovensko-maďarské (-0,052) a kontrolné školy s vyučovacím jazykom maďarským (-0,53). Významný rozdiel v priemerných matematických kompetenciách je len medzi školami kontrolnej skupiny s vyučovacím jazykom maďarským a ostatnými školami. Tento výsledok je zhodný s výsledkom získaným klasickou teóriou testov v [6]. Matematické kompetencie žiakov pri rozklade na všetky skupiny zobrazujú kategorizované škatuľové grafy na obrázku 5.



Obrázok 5: Kategorizované škatuľové grafy

4. Záver

Pri hodnotení výstupného testu projektu KEGA 3/7001/09 metódami IRT sme zistili, že všetky úlohy sú približne rovnakej náročnosti. Žiadalo by sa, aby náročnosť úloh testu mala väčšiu variabilitu. Najmenej vhodnou úlohou je úloha *Návšteva divadla*, ktorá poskytuje najmenej informácií a je nevhodne bodovaná. Pri ďalšom použití tohto testu by ju bolo treba vhodne upraviť. Najnáročnejšími úlohami sú úlohy z oblasti pravdepodobnosť (*Gulôčky*) a z oblasti logika (*Požičovňa bicyklov*).

Literatúra

- [1] Baker F.-Kim S. H. (2004). Item Response Theory: Parameter estimation techniques (2nd ed). New York:, Marcel Dekker. 2004.s. 528. ISBN 978-08-2475-825-0
- [2] Chernyshenko, O. S. and others (2001): Fitting Item Response Theory Models to Two Personality, Inventories: Issues and Insights. Multivariate Behavioral Research, 36 (4), 2001,s. 523-562. ISSN 0027-3171
- [3] Kóšová M.-Rybanský Ľ. (2011): Štatistické spracovanie výsledkov výstupného testu pre 6. ročník projektu KEGA 3/7001/09. In: Zborník príspevkov z IX. Nitrianskej matematickej konferencie organizovanej Katedrou matematiky FPV UKF v Nitre (22. –23. September 2011), Fakulta prírodných vied UKF v Nitre 2011. s. 117-123. ISBN 978-80-8094-958-7

- [4] Rizopoulos D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. Journal of Statistical Software, 17 (5), 2006. s. 1-25. ISSN 1548-7660. URL <http://www.jstatsoft.org/v17/i05/>
- [5] Rybanský Ľ.-Vrábelová M. (2010). Štatistické spracovanie výsledkov vstupného testu KEGA 3/7001/09. In: Zborník príspevkov z vedeckej konferencie Pedagogická veda a školská prax v historickom kontexte (28. Január 2010), Katedra pedagogiky Filozofickej fakulty Univerzity sv. Cyrila a Metoda v Trnave, 2010. s. 175-184. ISBN 978-80-8105-182-1
- [6] Rybanský Ľ.-Vrábelová M. (2010). Štatistické spracovanie výsledkov výstupného testu pre 5. ročník projektu KEGA 3/7001/09. In: Zborník príspevkov z VIII. Nitrianskej matematickej konferencie organizovanej Katedrou matematiky FPV UKF v Nitre (16. – 17. September 2010), Fakulta prírodných vied UKF v Nitre 2010. s. 225-232. ISBN 978-80-8094-781-1
- [7] Rybanský Ľ.-Vrábelová M. (2012). Využitie programového balíka ltm na spracovanie výsledkov projektu KEGA. In: Forum Statisticum Slovaca, č. 3, 2012. s. 126-131. ISSN 1336-7420
- [8] <http://info.worldbank.org/etools/docs/library/117765/Item%20Response%20Theory%20-%20F%20Baker.pdf>
- [9] <http://personality-project.org/r/book/Chapter8.pdf>
- [10] <http://www.metheval.uni-jena.de/irt/VisualIRT.pdf>

Článok prijatý dňa 27. júna 2012

Adresa autorov

*RNDr. Ľubomír Rybanský,
Katedra matematiky
Fakulta prírodných vied
Univerzita Konštantína Filozofa v Nitre
Tr. A. Hlinku 1
SK – 94974Nitra
e-mail: lubomir.rybansky@ukf.sk*

*doc.RNDr. Marta Vrábelová, CSc.,
Katedra matematiky
Fakulta prírodných vied
Univerzita Konštantína Filozofa v Nitre
Tr. A. Hlinku 1
SK – 94974Nitra
e-mail: mvrabelova@ukf.sk*

PodĎakovanie

Príspevok vznikol v rámci riešenia projektu KEGA 015 UKF - 4/2012.