

ПРИМЈЕНА ТЕОРИЈЕ ОДГОВОРА НА ПИТАЊА У ЕКСТЕРНОМ ОЦЈЕЊИВАЊУ УЧЕНИЧКИХ ПОСТИГНУЋА

Проф. др Стеван Р. Стевић
Економски факултет у Брчком
Душан Сарајлић, дипл. инж.
Агенција за стандарде и оцјењивање у образовању

Summary

Item Response Theory (IRT) is a modern statistic method allowing both the item analysis and the test results processing and analysing. The Rasch model and OPLM analysis make use of IRT methodology in practice in a range of analyses considering neither the examinee sample nor the observed distribution of their abilities.

The paper presents the IRT theory and its use for Mother Tongue and Mathematics 8th grade test items examination. For this purpose exploited were results of the external evaluation of pupil achievement in two core subjects on a sample of 3 143 8th grade pupils from 105 schools across BiH. Stated methodology has been aimed at setting up the standards of learning achievement and designing the Mother Tongue and Mathematics items database.

The Item Response Theory appeared to be efficient for testing the item validity and correcting the expert standards by empirical indicators observed from student achievement. Aside, the mentioned methodology can expose the weight of learning areas within the learning achievement output on the whole.

УВОД

Да би се предузимале адекватне мјере за унапређење квалитета образовања, као једног од приоритетних циљева реформе образовања у Босни и Херцеговини, потребно је, између осталог, обезбиједити валидне показатеље ученичких постигнућа и фактора који битно утичу на квалитет образовања у цјелини. Екстерно оцјењивање, као други поглед на квалитет образовног процеса, за разлику од интерног оцјењивања у школама, је један од начина оцјењивања који може да задовољи наведене потребе. Агенција за стандарде и оцјењивање у образовању за Федерацију Босне и Херцеговине и Републику Српску реализује пројекат Екстерно оцјењивање ученичких постигнућа у основној школи. Основни циљ екстерног оцјењивања је мјерење ученичких постигнућа из матерњег језика и математике на крају разредне наставе и на завршетку основне школе, као и утврђивање фактора који утичу на ниво образовних постигнућа ученика.

Средином маја 2003. године Агенција је провела екстерно оцјењивање ученичких постигнућа из матерњег језика и математике. Тестирано је 2 857 ученика четвртог разреда и 3 143 ученика осмог разреда ¹⁾, на узорку од 6 000 ученика, из 105 основних школа у Босни и Херцеговини. У дијелу пројекта који је обухватио ученике осмог разреда, у припреми испитних задатака пошло се од процјене експерата о томе шта би ученици требали да знају из наведених предмета (коришћени су експертски стандарди). На тај начин је припремљена полазна база испитних задатака из које су дизајнирани тестови за ученике. Анализом одговора на испитне задатке, добијени су резултати који за овај дио пројекта имају двоструки значај. Прије свега, на основу њих постављени су дефинитивни стандарди за матерњи језик и математику за осми/девети разред основне школе, што је основни предуслов за припрему Опште оцјене ученичких постигнућа, односно тестирања укупне популације ученика осмог/деветог разреда из наведених предмета. Поред тога, приступило се формирању базе испитних задатака из матерњег језика и математике за ученике осмог/деветог разреда, чија је функционалност практично провјерена и на тесту и у статистичкој анализи резултата тестирања. Дефинитивна база испитних задатака садржи само оне задатке који су се показали као функционални, односно који се уклапају у одговарајући статистички модел.

¹⁾ Више о томе видјети у Сиручном извјештају: Екстерно оцјењивање ученичких постигнућа у основној школи, Агенција за стандарде и оцјењивање у образовању за ФБиХ и РС, Сарајево, 2004. године

1. ОСНОВЕ ТЕОРИЈЕ ОДГОВОРА НА ПИТАЊА

Анализа теста и испитних задатака била је базирана на Теорији одговора на питања (IRT- **Item Response Theory**). За обраду података коришћен је низ програмских пакета, а најзначајнији за област статистичке анализе је OPLM (Verhelst et al., 1991). OPLM је софтверски програм који користи једнопараметарски Рашов модел («Rasch-ов модел») за анализу по IRT методологији.

Важна одлика психометријских модела базираних на Теорији одговора на питања је да се процјена нивоа тежине испитној задатка и испитаникове способности мјере истом јединицом мјере и да се могу представити на истој скали. Резултат је да садржај испитних задатака из теста даје значење, у односу на представљени критеријум, индивидуалним резултатима теста израженим на тој скали. Рашов модел је IRT модел и базира се на предвиђеној специфичној објективности и довољној стабилности необрађених резултата. Ово је могуће због тога што је процјена испитаникове способности независна у односу на даће испитне задатке. Другим ријечима, те процјене односе се на личне способности ученика обухваћене различитим испитним задацима, а резултат различитих сејова испитних задатака може бити изражен на истој скали, под условом да је уопријебљен повезани (преклајајући) дизајн теста.

У Рашовом моделу (Rasch, 1960.) вјероватноћа (P) правилног одговора ($x=1$) даћа је у следећој формули:

$$P\{x = 1 | \theta_v; \delta_i\} = \frac{e^{(\theta_v - \delta_i)}}{1 + e^{(\theta_v - \delta_i)}}$$

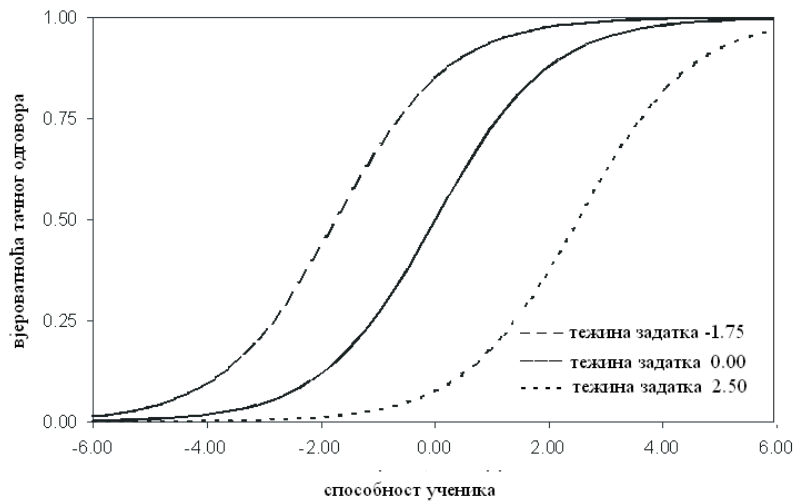
гдје је θ_v способност испитаника v , а δ_i је тежина испитног задатка i (неки статистичари користе римско слово B или b да би означили тежину испитног задатка).

Модел уписује вјероватноћу правилног одговора као функцију разлике између испитаникове способности и тежине испитног задатка. Ова вјероватноћа биће 0.5 ако испитаникова способност у потпуности одговара тежини испитног задатка. Вјероватноћа ће биће већа од 0.5 ако испитаникова способност прелази тежину испитног задатка, а биће мања од 0.5 ако тежина испитног задатка превазилази испитаникову способност.

Рашов модел и његова природна екстензија ка политомним подацима (могућност давања дјелимичних резултата) има довољно статистике за параметре способности, што је у складу са уобичајеном праксом у образовном оцјењивању, када се користи збирна оцјена. Модел има пожељне математичке одлике. То што обезбјеђује довољно статистике за параметре способности олакшава условну максималну вјероватноћу процјене параметара испитних задатака, што је предност јер не треба правити никакве претпоставке о дистрибуцији способности. Такође, са становишта процјене параметара, није потребно правити узорак испитаника методом случајног узорка. Друга погодност математичке структуре овог модела је могућност израде одговарајућих статистичких процедура за оцјену валидности модела за податке базиране на статистици са доказаном дистрибуцијом, које су корисне у случају специфичних повреда модела.

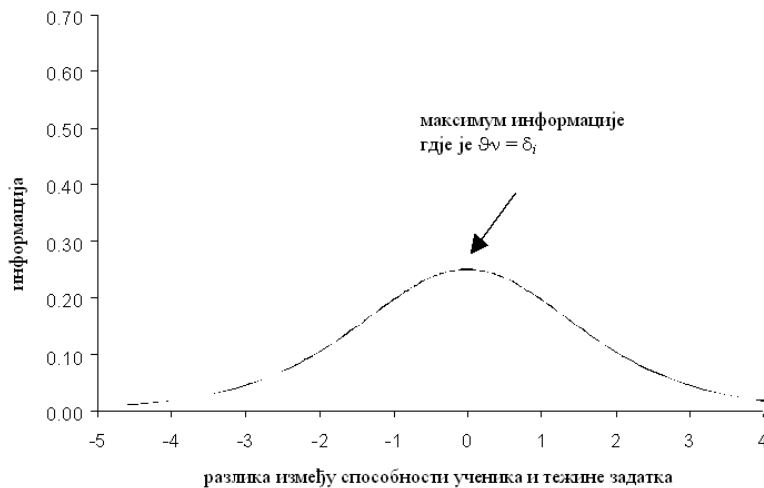
Однос између вјероватноће правилног одговора и функције разлике способности и тежине графички се може представити карактеристичном кривом испитног задатка, као што је приказано на слици 1.1. за три испитна задатка различите тежине.

Слика 1.1. Карактеристичне криве испитног задатка за три задатка различите тежине



Количина информација која треба да се добије од једног испитног задатка представља производ вјероватноће правилних и неправилних одговора. Тако информација расте како се тежина испитног задатка удаљава од испитаникове способности. Ако је испитни задатак јако тежак, неће се добити информација јер је мала вјероватноћа правилног одговора на такав испитни задатак. Такође, ако је испитни задатак јако лаган, ниво добијених информација је минималан јер је скоро сигурно (велика вјероватноћа) да ће бити тачан одговор. Функција информације је графички представљена на слици 1.2.

Слика 1.2. Функција информације испитног задатка



2. АНАЛИЗА ИСПИТНИХ ЗАДАТАКА ПРИМЈЕНОМ ТЕОРИЈЕ ОДГОВОРА НА ПИТАЊА

Обрада и статитичка анализа резултата тестирања, у којој су комбиновани експертски стандарди и показатељи ученичких постигнућа из матерњег језика и математике, послужила је за постављање дефинитивних стандарда ученичких постигнућа. Због тога је приликом статистичке анализе првенствено тестирана функционалност коришћених испитних задатака. У интерпретацији дати су почетни резултати и коначни резултати након елиминације нефункционалних испитних задатака.²⁾ У овом раду су, поред почетних и коначних резултата, наведене и разлике у тежинама испитних

²⁾ Више о томе видјети у Сврчучном извјештају: Екстерно оцјењивање ученичких постигнућа у основној школи, Агенција за стандарде и оцјењивање у образовању за ФБиХ и РС, Сарајево, 2004. године

задатака по полу ученика, као и примјери функционалних и нефункционалних испитних задатака и критеријуми за њихову елиминацију. У прелиминарној анализи резултата дате су и функције информација по појединим подручјима у оквиру сваког предмета, добијене након финалне анализе примјеном IRT методологије.

2.1. Резултати прве анализе

Резултати прве анализе, која је обухватила одговоре свих ученика на све испитне задатке, приказани су у Табели 2.1.1. за математику, а у Табели 2.1.2. за матерњи језик. Табеле показују број тест-књижице, број ученика по тест-књижици и нека статистичка мјерења. Око 300 ученичких одговора добијено је на испитне задатке у свакој тест-књижици, а свака тест-књижица имала је око 40 испитних задатака. За сваку тест-књижицу приказани су: средња вриједност резултата, стандардна девијација, поузданост теста (α), средња вриједност способности ученика (\mathcal{G}) и варијанса способности ученика (σ^2_0).

2.1.1. Математика

Процјене поузданости теста за математику биле су у распону од 0.81 до 0.88, што се може сматрати веома задовољавајућим за један прелиминарни тест. Вриједности приказане у Табели 2.1.1. показују да су тест-књижице биле упоредиве по средњој вриједности, стандардној девијацији и поузданости теста (што је била намјера). Средња вриједност способности ученика и варијанса биле су прилично уједначене по тест-књижицама, али је тежина испитних задатака била превисока за ученике који су имали пропорцију тачних одговора у просјеку мању од 0.25. Очигледно је да су очекивања о способности ученика осмог разреда била превисока.

Може се уочити да разлике у добијеним пропорцијама тачних одговора у овој табели садрже ”помијешане информације”, првенствено због тога што тест-књижице имају различите комплете испитних задатака и различите ученике, па ниска пропорција у одређеној тест-књижици може теоретски бити проузрокована посебно тешким комплетом испитних задатака, нарочито ниском средњом вриједношћу способности ученика, или комбинацијом оба ова узрока. Међутим, средња вриједност способности ученика (\mathcal{G}) у овим табелама узима у обзир разлику у тежини између тест-књижица и може бити директно упоређивана. Како се јасно види из Табеле 2.1.1. процјене способности су прилично близу очекиваним зато што су тест-књижице биле спирално распоређене међу тестираним школама.

Табела 2.1.1. Прва анализа свих испитних задатака из математике по тест-књижицама

Број тест-књижице	1	2	3	4	5	6	7	8	9	10
Број ученика	302	300	298	303	295	297	302	301	302	297
Број задатака	44	44	44	44	44	44	44	44	44	44
Средња вриједност резултата	10.35	11.06	12.81	9.50	7.97	10.98	9.78	9.70	10.35	9.62
Пропорција тачних одговора	0.24	0.25	0.29	0.22	0.18	0.25	0.22	0.22	0.24	0.22
Стандардна девијација	6.67	5.61	6.81	6.06	5.07	6.74	5.89	6.16	6.32	6.76
Поузданост теста (α)	0.87	0.81	0.85	0.84	0.82	0.88	0.84	0.85	0.84	0.86

Средња вриједност способности ученика (\bar{g})	-1.71	-1.95	-1.74	1.62	-1.76	-1.64	-1.72	-1.89	-1.88	-1.79
Варијанса способности ученика (σ^2_{θ})	1.13	0.82	0.90	0.91	0.93	1.29	1.09	1.10	0.82	0.89

2.1.2. Матерњи језик

Процјене поузданости теста за матерњи језик биле су у распону од 0.81 до 0.84, што се, као и код математике, може сматрати задовољавајућим за овакав тест. Вриједности дате у Табели 2.1.2. показују да су тест-књижице за матерњи језик, такође, биле прилично упоредиве, што се тиче средњих вриједности, стандардне девијације и поузданости резултата. Средња вриједност способности ученика и варијанса биле су уједначене по тест-књижицама.

Средња вриједност способности ученика (\bar{g}) протеже се од -0.43 до -0.01 , што показује да су испитни задаци из матерњег језика били ближе средњој вриједности способности ученика него испитни задаци из математике (пропорција тачних одговора је у просјеку била нешто преко 0.50).

Табела 2.1.2. Прва анализа свих испитних задатака из матерњег језика по тест-књижицама

Број тест-књижице	1	2	3	4	5	6	7	8	9	10
Број ученика	301	304	303	305	296	299	302	300	305	300
Број задатака	42	41	43	41	38	39	41	42	42	43
Средња вриједност	21.52	22.39	22.96	20.78	19.07	18.77	21.58	21.67	20.96	22.05
Пропорција тачних одговора	0.51	0.55	0.53	0.51	0.50	0.48	0.53	0.52	0.50	0.51
Стандардна девијација	7.06	6.54	6.60	6.72	6.43	6.26	6.44	6.86	6.95	6.31
Поузданост теста (α)	0.82	0.81	0.82	0.82	0.82	0.81	0.82	0.84	0.83	0.80
Средња вриједност способности ученика (\bar{g})	-0.02	-0.18	-0.15	-0.05	0.07	0.12	-0.01	-0.23	-0.43	-0.25
Варијанса способности ученика (σ^2_{θ})	0.69	0.69	0.77	0.73	0.83	0.70	0.74	0.85	0.79	0.59

2.2. Принципи и процедуре избора испитних задатака

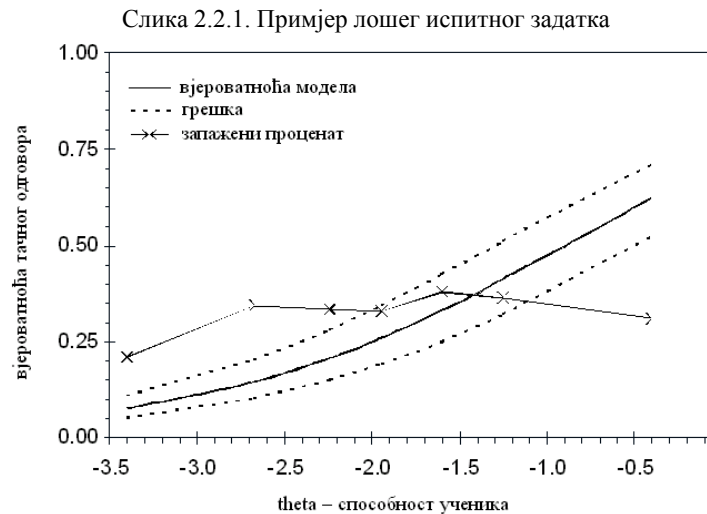
Након прве анализе испитних задатака за оба предмета, показало се да дискриминаторна снага и количина информација варира од задатка до задатка. Због тога је извјестан број задатака требало одстранити. Главни критеријум за одбацивање испитних задатака био је недостатак уклапања у модел у комбинацији са ниским дискриминаторним индексом. Осим нумеричких података графикони са емпиријском кривом испитних задатака обезбиједили су ефикасну процедуру за брзо провјеравање уклапања и дискриминаторности модела. У овом поглављу дати су подаци за два примјера испитних задатака: један који треба да буде одбачен и други који показује одлично уклапање у модел и који треба да буде задржан у бази испитних задатака.

2.2.1. Примјер лошег и доброг испитног задатка

Као примјер испитног задатка који је показао лош квалитет узели смо испитни задатак број 5. За овај испитни задатак статистику неслањања са моделом је изузетно висока, ($S=128.512$), због чега је вјероватноћа поклањања са моделом минимална ($P=0.000$).

nr	label	A	B	SE(B)	S	DF	P	M	M2	M3
5	J1920461	1	-0.900	0.097	128.512	6	0.000	8.100	8.495	9.9609

Индекси за правце девијације (M, M2 и M3) такође показују веома високе позитивне вриједности, указујући да испитни задатак не прави велику дискриминацију између ученика који су добро урадили остале испитне задатке из математике и оних ученика који су урадили слабије. Овај испитни задатак добија ознаку "лош" и искључује се из даље анализе. Слика 2.2.1. показује графичку илустрацију карактеристика испитног задатка.



Хоризонтална оса представља способност ученика, а вертикална вјероватноћу тачног одговора. Ако испитни задатак функционише како је пожељно, вјероватноћа тачног одговора ће се повећавати како се увећава способност. Ово очекивање је приказано правом линијом која се пружа из лијевог доњег угла графика према десном горњем углу. Та линија се назива «карактеристична крива испитног задатка». Одступање од модела унутар процијене грешке са становишта статистике сматра се неважним. Границе појаса грешке су означене паралелним тачкастим линијама са обје стране карактеристичне криве. Крстићи повезани правом линијом показују запажања, тј. проценат тачних одговора ученика на означеном нивоу способности. Када оне падају изван појаса грешке, испитни задатак представља озбиљан ризик или неуклапање. На четири од укупно седам нивоа способности добијени резултати за овај испитни задатак падају изван појаса грешке.

Поред тога, може се видјети да је крива запажених резултата прилично равна, што значи да су ученици различитих способности имали прилично сличан проценат правилних одговора за овај испитни задатак. Ова одлика се назива "ниска дискриминација", што значи да испитни задатак није правно разлику између способних и мање способних ученика.

Испитни задатак бр.16 из матерњег језика представља примјер доброг задатка. Резултати калибрације били су следећи:

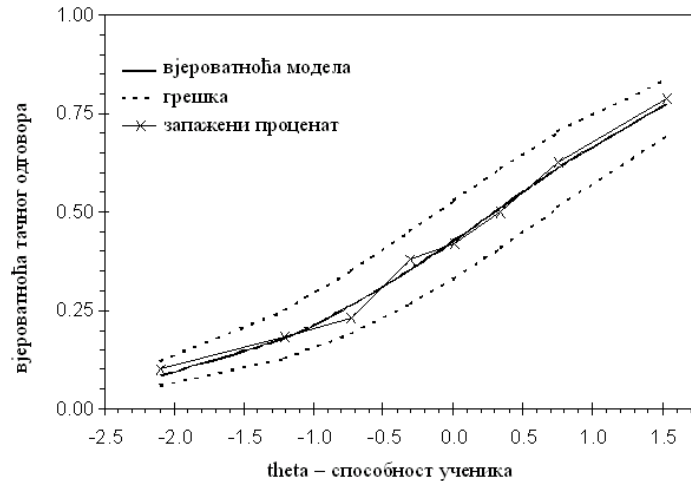
nr	label	A	B	SE(B)	S	DF	P	M	M2	M3
16	K3520294	1	0.335	0.100	0.890	7	0.996	0.358	0.033	-0.408

Статистика овог испитног задатка је доста ниска ($S=0.890$), а вјероватноћа поклањања са моделом износи $P=0.996$. Индекси за правац девијације (M, M2 и M3) су веома близу нуле, указујући да је

испитни задатак дискриминаторан, како је и очекивано. Због тога се овај испитни задатак задржава у даљој анализи. Слика 2.2.2. представља графикон овог задатка.

Слика 2.2.2. јасно показује да добијене вриједности (линија са крстићима) падају близу очекиваних вриједности (права линија). Овај испитни задатак се готово идеално поклапа са моделом.

Слика. 2.2.2. Примјер доброг испитног задатка



2.3. Резултати након селекције

Испитни задаци су изабрани помоћу више OPLM анализа према раније описаној процедури. После прве OPLM анализе избрисани су испитни задаци који се нису уклапали у модел и који су показали лошу дискриминацију. Као што се очекивало, у другој анализи неки испитни задаци су показали неодговарајуће вриједности, па су и они избачени. Ова процедура је настављена током 6-7 анализа, све док нису добијени задовољавајући показатељи уклапања у модел. Резултати коначне анализе представљени су у Табелама 2.3.1. и 2.3.2.

2.3.1. Математика

Селекција испитних задатака довела је до знатно умањене базе података испитних задатака из математике: скоро 50% испитних задатака су одбачени на бази критеријума психометријског квалитета. Међутим, преостали испитни задаци достигали су скоро исту поузданост, па је процес избора знатно повећао ефикасност. У ствари, RMS (средња вриједност квадратног коријена) поузданости у 10 тест-књижица била је 0.85. Под условом да се смањи просјечна дужина текст-књижице за 51%, Спирман-Броун процјена би дала RMS поузданост од 0.73. Путем процеса селекције RMS поузданост умањеног броја испитних задатака износи у ствари 0.81. Овај преостали сет испитних задатака је прилично доброг квалитета да би обезбиједио поуздана и вјеродостојна мјерења у наредним годинама. Пропорција тачних одговора по тест-књижици је достигла просјек од 0.32, што је прилично ниско, али одсликава стварну ситуацију што се тиче ученичких постигнућа из математике.

2.3.2. Матерњи језик

Селекција испитних задатака из матерњег језика довела је до смањења базе испитних задатака за трећину. Преостали сет испитних задатака је високог квалитета. У ствари, поузданост овог умањеног сета задатака је апсолутно већа од поузданости почетне групе испитних задатака. RMS поузданост у 10 тест-књижица је била 0.82. Под условом да се умањи просјечна дужина тест-књижице за 35%, Spearman-Brown процјене би дале RMS поузданост од 0.75. Процесом избора RMS поузданост смањене групе испитних задатака износи 0.83 и омогућиће ефикасно, поуздано и валидно мјерење у будућности.

Табела 2.3.1. Подаци за математику по тест-књижицама, одабрани испитни задаци (коначна анализа)

Број тест-књижице	1	2	3	4	5	6	7	8	9	10
Број ученика	244	241	259	228	208	240	230	221	250	249
Број задатака	20	22	19	20	21	20	22	25	25	20
Средња вриједност резултата	6.86	7.19	7.19	5.58	5.40	6.83	7.02	8.37	7.35	5.90
Пропорција правилних одговора	0.34	0.33	0.38	0.28	0.26	0.34	0.32	0.33	0.29	0.30
Стандардна девијација	4.35	3.70	3.88	4.19	3.57	4.09	4.19	4.81	4.59	4.22
Поузданост теста (α)	0.83	0.76	0.79	0.83	0.79	0.83	0.82	0.82	0.81	0.81
Средња вриједност способности ученика (\mathcal{G})	-1.09	-1.41	-1.17	-0.77	-0.92	-1.02	-1.09	-1.26	-1.45	-1.29
Варијанса (σ^2_{θ})	1.68	1.42	1.47	1.80	1.62	2.07	1.88	1.49	1.36	1.40

2.4. Прелиминарна интерпретација резултата тестирања

У наредном поглављу дато је прелиминарно тумачење резултата тестирања добијених на основу базе података испитних задатака за осми разред из математике и матерњег језика. Треба схватити да су ово прелиминарне оцјене, јер основни циљ тестирања нису била ученичка постигнућа, него конструисање одговарајуће банке испитних задатака за употребу у наредном периоду. Због тога су и информације о ученичким постигнућима које се базирају на овим резултатима мање сигурне и поуздане. Међутим, резултати тестирања ученика осмог разреда могу бити употријебљени да би се формулисале хипотезе о ученичким постигнућима и могућим стандардима које треба оцијенити коришћењем одабране базе података испитних задатака у наредним годинама.

2.4.1. Анализа по подручјима: математика и матерњи језик – функције информација

Слика 2.4.1. приказује функције информација испитних задатака за шест испитних подручја у бази података испитних задатака из математике за осми разред. Да би се омогућило упоређивање за различита испитна подручја, информационе функције су подијељене на нивое, како би представиле једнаку величину комплета испитних задатака. Хоризонтална права представља способност ученика, а вертикална количину информација добијену из одабраних комплета испитних задатака за свако испитно подручје.

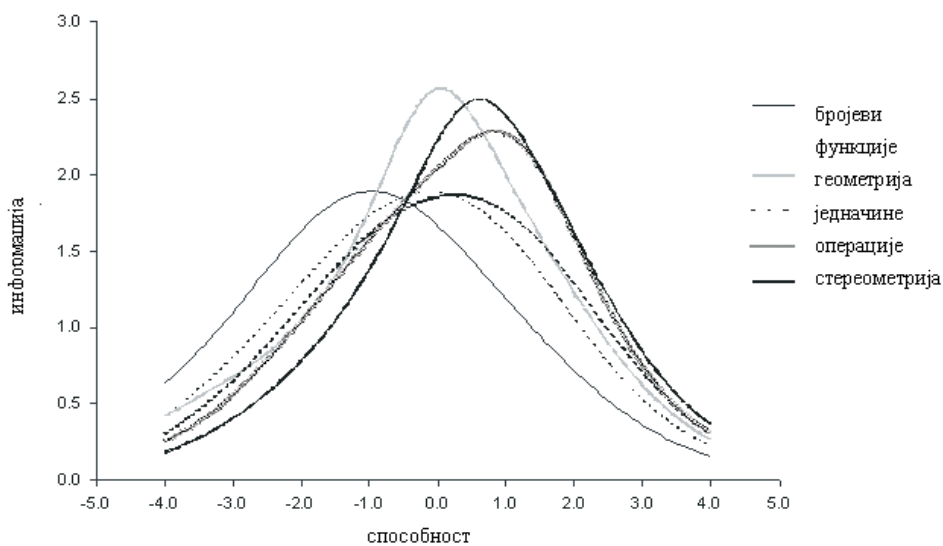
Функција информације илуструје такозвану 'локалну поузданост', тј. количина информација добијена из комплета испитних задатака није иста на свим нивоима способности. Слика 2.4.1. приказује да испитни задаци у подручју Бројеви обезбјеђују скоро исту количину информација као и испитни задаци из подручја Функције и Једначине. Међутим, испитни задаци из подручја Бројеви су прикладнији за нижи ниво способности у осмом разреду од испитних задатака из подручја Функције и Једначине. Испитни задаци из подручја Геометрија достижу врхунац информације на средини распона способности, док испитни задаци из подручја Геометрија и Стереометрија пружају више информација што се иде ка врху скале.

Релативне функције информације за комплет испитних задатака за четири испитна подручја у банци података испитних задатака из матерњег језика за осми разред приказане су на слици 2.4.2., гдје се јасно види да испитни задаци из подручја Култура писменог изражавања обезбјеђују знатно више информација на свим нивоима способности, осим најнижег. Испитни задаци из других подручја не разликују се много по количини информација које обезбјеђују, иако су испитни задаци из Граматике подеснији на вишим нивоима, а испитни задаци из Књижевности функционишу боље од задатака из свих осталих подручја на најнижем нивоу способности.

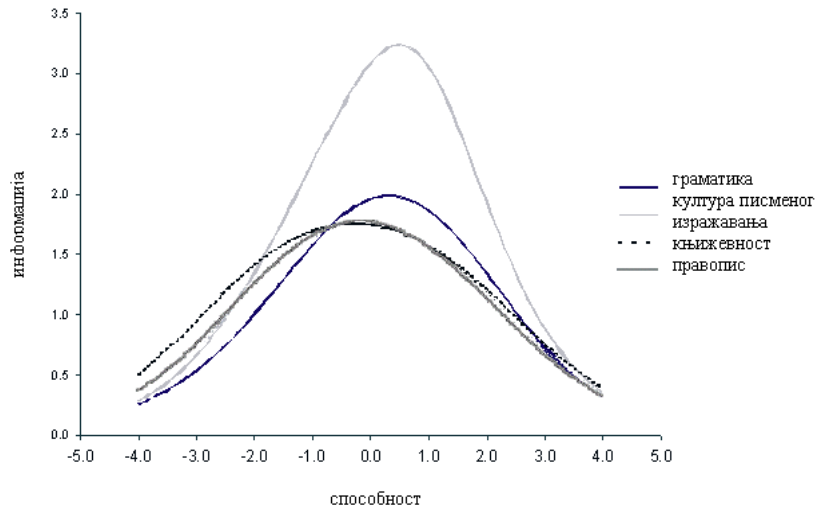
Табела 2.3.2. Подаци за матерњи језик по тест-књижицама, одабрани испитни задаци (коначна анализа)

Број тест књижице	1	2	3	4	5	6	7	8	9	10
Број ученика	301	304	303	305	296	299	302	300	305	300
Број задатака	26	29	29	27	27	27	29	29	23	22
Средња вриједност резултата	14.48	17.44	15.99	13.27	13.64	13.10	16.10	15.23	11.23	11.65
Пропорција правилних одговора	0.56	0.60	0.55	0.49	0.51	0.49	0.56	0.53	0.49	0.53
Стандардна девијација	5.87	5.74	5.51	5.30	5.09	4.94	5.48	5.73	4.88	4.41
Поузданост теста (α)	0.84	0.83	0.84	0.83	0.82	0.81	0.81	0.84	0.84	0.80
Средња вриједност способности ученика (\mathcal{G})	-0.04	-0.20	-0.18	-0.09	0.19	0.24	0.23	-0.05	-0.46	-0.21
Варијанса способности ученика (σ^2_{θ})	1.25	1.09	1.32	1.30	1.35	1.12	0.99	1.20	1.70	1.22

Слика 2.4.1. Функције информације по испитним подручјима за математику



Слика 2.4.2. Функције информације по испитним подручјима за матерњи језик

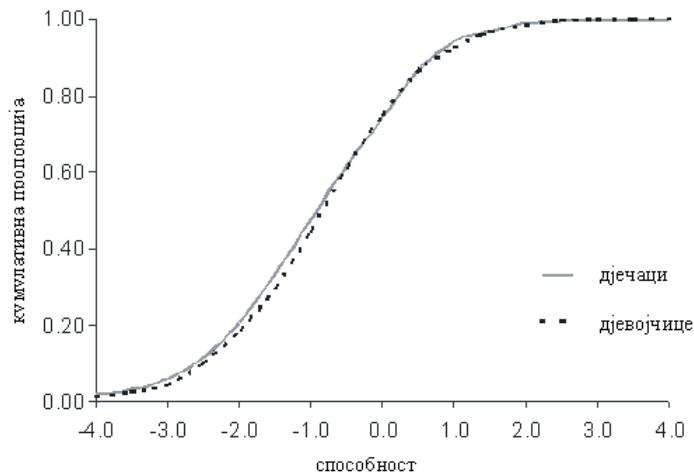


2.4.2.1. Разлике по полу за математику и матерњи језик

Слика 2.4.2.1. показује кумулативну дистрибуцију функција за процјене способности по полу добијене из одабраног сета задатака за математику. Хоризонтална оса представља способност ученика, а вертикална оса кумулативну пропорцију дјечака и дјевојчица осмог разреда.

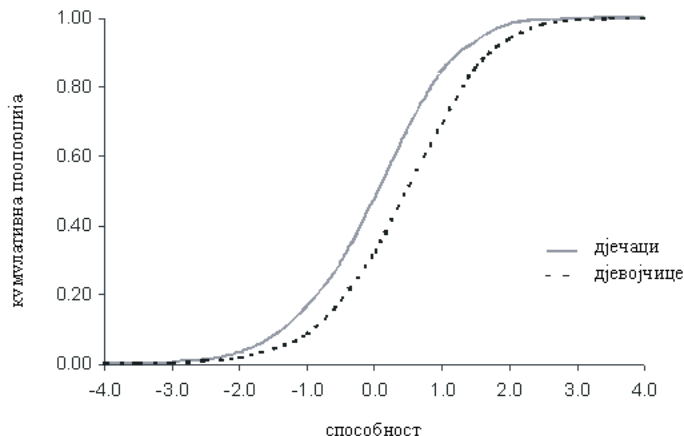
Наведена слика показује (да је процијењено) да само 5% и једног и другог пола има способност из математике испод -3.0 на скали, а да само око 5% има способност изнад 1.0. Међутим, најинтересантија информација из ове слике је да дјевојчице и дјечаци имају скоро исту дистрибуцију, премда дјевојчице показују мало боље резултате. Ово је нови феномен у образовању, јер су традиционално дјечаци бољи у математици.

Слика 2.4.2.1. Кумулативна дистрибуција по полу - математика



Слика 2.4.2.2. показује кумулативну дистрибуцију функција за процијењене способности по полу, добијене из одабраног сета задатака за матерњи језик. Овдје се може видјети да дјевојчице имају прилично боље резултате од дјечака.

Слика 2.4.2.2. Кумулативна дистрибуција по полу- матерњи језик



ЗАКЉУЧАК

Теорија одговора на питања спада у савремене статистичке методе које омогућавају не само анализу испитних задатака него и обраду и анализу резултата оцјењивања ученика (или других испитаника). Рашов модел и OPLM анализа конкретизују примјену IRT методологије за различите анализе, без обзира на изабране узорке испитаника и добијене дистрибуције њихових способности.

У раду су дате теоријске основе IRT методологије и њена примјена у анализи испитних задатака из матерњег језика и математике за осми разред основне школе. Коришћени су резултати екстерног оцјењивања ученичких постигнућа из наведених кључних предмета, на узорку од 3 143 ученика осмог разреда из 105 основних школа у БиХ. Наведена методологија примијењена је с циљем постављања стандарда образовних постигнућа и формирања базе испитних задатака из матерњег језика и математике на завршетку основне школе.

Показано је да се Теорија одговора на питања може ефикасно користити за провјеру ваљаности испитних задатака и корекцију експертских стандарда емпиријским показатељима ученичких постигнућа. Поред тога, примјеном наведене методологије може се утврдити и значај појединих испитних подручја у укупној информацији о образовним постигнућима ученика из одговарајућег наставног предмета.

ЛИТЕРАТУРА:

1. Baker, Frank (2001) The Basics of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD. (Available on Internet: <http://edres.org/irt/baker/>)
2. Kirsch, I., De Jong, J.H.A.L., Lafontaine, D., McQueen, J., Mendelovits, J., and C. Monseur. (2002) Reading for Change: Performance and Engagement Across Countries - Results from PISA 2000. Paris: OECD.
3. Rasch, G. (1960) Probabilistic Models for Some Intelligence and Attainment Tests. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.
4. Стручни извјештај (2004) Екстерно оцјењивање ученичких постигнућа у основној школи, Агенција за стандарде и оцјењивање у образовању за ФБиХ и РС, Сарајево.
5. Стевић, С., Петровић, Ј. (2004) Образовна постигнућа ученика четвртог разреда основне школе у БиХ, Настава, Републички педагошки завод, Бања Лука.
6. Verhelst, N.D., Glas, C.A.W., and H.H.F.M. Verstralen (1991) OPLM: A Computer Program and Manual. Arnhem: CITO.