

**Софийски университет „Св. Климент Охридски”
Факултет по математика и информатика**



Павлин Иванов Цонев

**МЕТОДИ ЗА АНАЛИЗ НА РЕЗУЛТАТИ ОТ
СТАНДАРТИЗИРАНИ ОЦЕНЯВАНИЯ НА УЧЕНИЦИ ПО
МАТЕМАТИКА**

**Автореферат
на дисертационен труд**

**за придобиване на образователната и научна степен „доктор“
в област на висше образование 1. Педагогически науки,
професионално направление 1.3. Педагогика на обучението по...**

**Докторска програма
„Методика на обучението по математика и информатика“
Методика на обучението по математика**

**Научен ръководител
проф. Кирил Банков**

София, 2025

Съдържание

Увод	2
ПЪРВА ГЛАВА. Анализ на НВО по математика след VII клас чрез КТТ.....	6
Класическа теория на тестовете – основни положения.....	6
Анализ на задачи от НВО VII клас	6
Обобщения и изводи	8
ВТОРА ГЛАВА. Анализ на НВО по математика след VII клас чрез IRT	11
Основни принципи и модели	11
Избор на подходящ модел за изследване на задачите с избираем отговор	16
Сравняване на резултатите на задачите с ИО за периода 2020-2023 г.	22
Точкуване на задачите със СО	23
ТРЕТА ГЛАВА. Експеримент за апробация на тест за НВО по математика след VII клас	29
Постановка на задачата.....	29
Анализ на получените резултати след провеждане на експеримента.....	33
Сравняване по различни показатели на традиционния и експерименталния начин на точкуване.....	36
Изводи от Трета глава.....	38
Заклучение	38
Приноси на дисертационния труд	42
Насоки за бъдещо развитие	44
Литература	44
Публикации на автора, свързани с темата на дисертацията	47
Декларация за оригиналност	48
Благодарности.....	49

Увод

Актуалност на темата

Оценяването на учениците е основен елемент в процеса на тяхното образование. Оценяването в България бива текущо или заключително, като последното може да е срочно, годишно и при завършване на даден етап на образование. През 2007 година Министерството на образованието и науката (МОН) въведе Национално външно оценяване (НВО) за учениците, завършващи IV, VII и X клас, и Държавен зрелостен изпит (ДЗИ) за завършващите XII клас. Предметът Български език и литература (БЕЛ) е задължителен за всички НВО и ДЗИ. Математиката е задължителна за НВО, докато за ДЗИ е по желание. За периода от 2007 до 2024 година НВО по математика след VII клас е съдържало различен брой задачи от трите типа – с избираем отговор (ИО), с кратък свободен отговор (СО) и разширен СО. В таблица 1 е даден броят на задачите по вид за всяка година. Както се вижда, само първите три години задачите са били единствено с ИО. В следващите години те са разделени в два модула, като в първи модул основно са задачите с ИО, а във втори – тези със СО.

Години	Общо задачи	I-ви модул		II-ри модул	
		ИО	Кратък СО	Кратък СО	Разширен СО
2007-2009	50	50	Няма модули		
2010-2011	30	25	-	3	2
2012-2018	24	16	4	2	2
2019	25	17	3	2	3
2020-2023	23	18	2	-	3
2024	23	20	-	-	3

Таблица 1. Брой задачи по вид за периода 2006-2024 г.

Всяка година, след приключването на външните оценявания се правят класирания и анализи на постигнатите резултати. Често тези анализи са основани на средните стойности на получените балове. Най-често се сравняват тези балове между основните предмети БЕЛ и Математика. Също така, на базата на тези балове, се правят изводи за нивото на успеваемост в различните административни области на страната, между населените места в тях, между различните училища в тези населени места. Резултатите от НВО след VII клас служат за класиране в профилирани и професионални паралелки в гимназиите. Доколко това е целесъобразно може да се проследи в (Банков, 2024). Този въпрос не е основна тема на настоящия дисертационен труд, макар че отчасти ще той бъде дискутиран.

В някои случаи анализирането на резултатите от външните оценявания се прави чрез използване на научни теории за изследване на тестовете. Най-често за тази цел се прилага добре известната Класическа теория на тестовете (КТТ) – например (Данчев, П., К. Банков, В. Стоименова и др., 2015), (Джалев, 2013), (Джалев, 2014) и пр. В световен план за начало на КТТ се счита книгата на Торндайк от 1904 г. (Thorndike, 1904). В България тази теория е представена от няколко автори – (Бижков, 1996), (Стойнова, 1996), (Банков, 2012).

Използването на КТТ, въпреки нейните отлични качества, има известни ограничения. В края на XX век в научната литература се появява нова теория, която се базира на вероятностните методи. Тя е известна като Item Response Theory (IRT), като през 1960 г. Г. Раш я описва за пръв път. В България това се случва през 1996 г. от Ф. Стойнова като част от книгата (Бижков, 1996). Вероятностното моделиране се наблюдава и при широкомащабни международни изследвания, като TIMSS (Trends in Mathematics and Science Study) и PISA (Programme for International Student Assessment). Повече за предимствата на IRT, за нейните принципи и развитие във времето, както и за някои от слабостите на КТТ ще стане дума във втора глава на дисертационния труд.

От самото начало на въвеждането на НВО в края на VII клас възникнаха редица въпроси, като:

- Как да се тълкуват резултатите на учениците?
- Покриват ли учениците в VII клас държавните образователни изисквания според тези резултати?
- Задачите, които се дават на НВО дали отговарят на някакви предварителни изисквания към тях?
- Използват ли се получените данни от НВО за да се направят изводи за подобряване на обучението по математика?
- Добре ли е тези резултати да се използват за вход в гимназиите?

В търсенето на отговори по част от тези въпроси, в настоящия дисертационен труд се анализират резултатите на учениците от НВО след VII клас, проведени в периода 2020-2023 година в четири административни области на България – София-град, Пловдив, Плевен и Разград. Направена е съпоставка между изводите, получени според двете

теории на тестовите. Разгледани са техните предимства и недостатъци. Направените изводи са използвани за провеждане на педагогически експеримент с ученици, завършващи VII клас през 2024 г., който е описан в трета глава. Там се изследва хипотезата, че вероятностното моделиране може да се използва за по-прецизно изследване на качествата на тестовите задачи и чрез него могат да се въведе по-опростено и по-обективно оценяване на учениците. В разработката на този труд е използвана методологията, описана в книгата (Иванов, 2006).

Поставените по-горе въпроси, както и други проблеми, свързани с НВО редовно се дискутират на Пролетните конференции на Съюза на математиците в България, на които авторът на дисертационния труд редовно взема участие. Една от първите такива дискусии, след започването на мащабните национални външни оценявания, се проведе през 2010 г (Банков К., Витанов Т., 2010). Самите теми се възлагат от МОН на Центъра за оценяване в предучилищното и училищното образование (ЦОПУО, 2024), като за подготовката им се привличат учители от различни училища от цялата страна. От 2010 г. авторът на този труд редовно участва като член на експертните комисии за съставяне на тестовите задачи по математика в ЦОПУО (тогава с абревиатурата ЦКОКО). Също така, от самото начало на провеждането на НВО, той е член и на регионалните комисии за проверката и оценяването им.

Цел и задачи на изследването

Основна цел в дисертационния труд е да се приложат различни методи за анализ на резултати от стандартизирани оценявания на ученици по математика (НВО), да се анализират положителните и отрицателните страни на всеки от тях и да се дадат препоръки за тяхното използване.

За постигането на тази цел в дисертационния труд се решават следните *основни задачи*:

- 1) Да се направи анализ на научно-теоретичните въпроси по темата на дисертационния труд, които очертават проблемната област на разглежданата тема;
- 2) Да се направи анализ на учебната документация, свързана с обучението по математика в прогимназиалния етап за придобиване на основно образование;
- 3) Да се направи анализ на тестовите задачи от НВО за периода 2020-2023 г. чрез КТТ;

- 4) Да се направи анализ на тестовите задачи от НВО за периода 2020-2023 г. чрез IRT;
- 5) Да се сравнят получените резултати по различните методи;
- 6) Да се извърши педагогически експеримент с ученици, които им предстои да завършат VII клас през 2024 г., който да използва направените изводи от проведените изследвания;
- 7) Да се предложат варианти за пресмятане на състезателния бал на учениците.

Обект на изследването са тестовете за НВО и методите за анализиране на резултатите от тях.

Предмет на изследването: Изпитни тестове за НВО в периода 2020-2023 г. и експериментална група с ученици в VII клас от различни училища в няколко населени места в България от периода април-юни 2024.

Методи на изследването

Методите, които се прилагат са: проучване на документи (включително тестови задачи и тяхната класификация); методи за анализи на резултати от тестове чрез КТТ и IRT; изводи от тези анализи и сравняването им; статистическа обработка на данните от проведените НВО; графично представяне на резултатите.

Хипотеза на изследването

Правилното използване на КТТ съчетано със съвременни методи за анализ на тестови резултати (IRT) дава по-обективна, по-точна и надеждна картина за постиженията по математика на изследваната популация.

Структура на дисертационния труд

Дисертационния труд съдържа увод, три глави, заключение, списък с използваната литература и приложения. Основната част на дисертационния труд е от 177 страници, в които са включени 105 фигури и 104 таблици. Приложенията са 8. Литературата съдържа 42 заглавия. Фигурите в дисертационния труд са отбелязани в средата под самата фигура, като е използвано означението **Фиг. Г.№.**, където с Г е означен номера на главата, в която е дадената фигура, а № е номерът на самата фигура в тази глава. Таблиците са отбелязани

под самата таблица с означението **Таблица Г.№.**, където Г е номерът на главата, а № е номерът на съответната таблица в главата. Цитирането на литературата е в стил АРА.

ПЪРВА ГЛАВА. Анализ на НВО по математика след VII клас чрез КТТ

Класическа теория на тестовите – основни положения

В първа точка на тази глава се дават кратки сведения за КТТ, някои исторически бележки, свързани с нея и се разясняват понятията, използвани в тази теория. След това се представя тестовата спецификация на НВО по математика след VII клас. Тя съдържа две основни направления – съдържателно и познавателно. Съдържателното направление се описва с различните теми от учебното съдържание, а познавателното – с мисловните дейности, които са необходими за решаването на задачите. Авторите на тестовите задачи трябва да ги съставят така, че те да са от различни съдържателни и познавателни области.

Измерваните елементи от съдържателните области са определени въз основа на учебното съдържание по математика, включено в учебните програми за пети, шести и седми клас. Тези области са разпределени в 27 теми, които са групирани в четири направления: 1) Числа, алгебра; 2) Фигури и тела. Измерване; 3) Елементи от вероятности и статистика; 4) Логически знания. Моделиране.

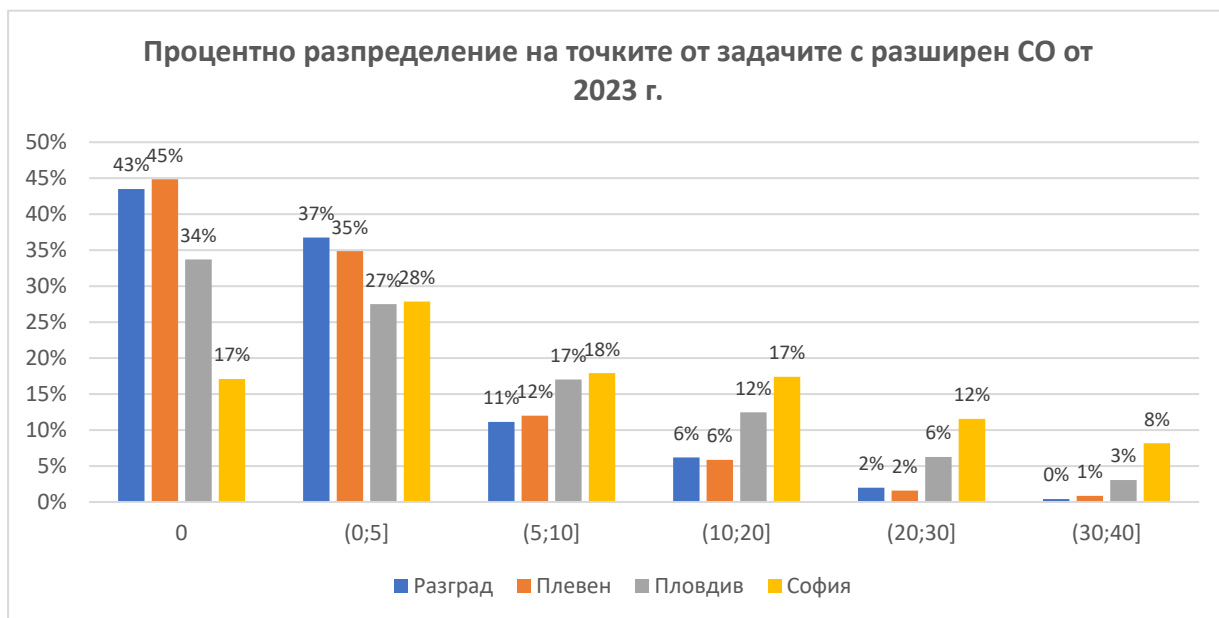
Освен съдържателната област всяка тестова задача се класифицира и по друг параметър – познавателна (когнитивна) област (Банков, 2012). Познавателните области описват видовете мисловна дейност, които са необходими за решаване на задачите от теста. Според известната таксономия на Блум тези познавателни равнища са шест: познаване (знание), разбиране, приложение, анализ, синтез и преценяване. В приложение 1 на дисертационния труд е показана текстовата спецификация, с която трябва да се съобразяват авторите на задачите за НВО по математика след VII клас през учебната 2018/2019 г. За следващите години тази спецификация не се различава съществено.

Анализ на задачи от НВО VII клас

В следващите четири точки подробно са разгледани НВО по математика след VII клас съответно за 2020, 2021, 2022 и 2023 година. Използвани са данни за резултатите на учениците от 4 области на България: София-град, Пловдив, Плевен и Разград. За всяка от тези четири години:

- се дават основните статистически характеристики на задачите с ИО, получени според КТТ;

- изобразява се графично разпределението на учениците според суровите им балове, като са представени няколко варианта за трансформацията на тези балове в оценки по общоприетата в България шестобална система;
- чрез критерия χ^2 на Пирсън е проверена хипотезата за нормалност на тези балове, като хипотезата H_0 е отхвърлена, т. е. не може да се твърди, че генералната съвкупност от баловете на учениците е нормално разпределена;
- задачите с ИО са подредени според коефициентите на трудност и на разграничителна сила и са представени таблици, характеризиращи дистракторите на всяка задача;
- на тази база са разгледани някои конкретни задачи от всяко от изследваните НВО, които са показали някои недотам добри характеристики по горните показатели. Обърнато е внимание на съдържателната област на всяка задача; на познавателното равнище, което тя покрива; потърсени са причините за по-лошите показатели, които са се проявили; дадени са варианти за подобряване на тези показатели; анализирано е представянето на учениците от различните административни области;
- коментирани са надеждността и валидността на тестовете;



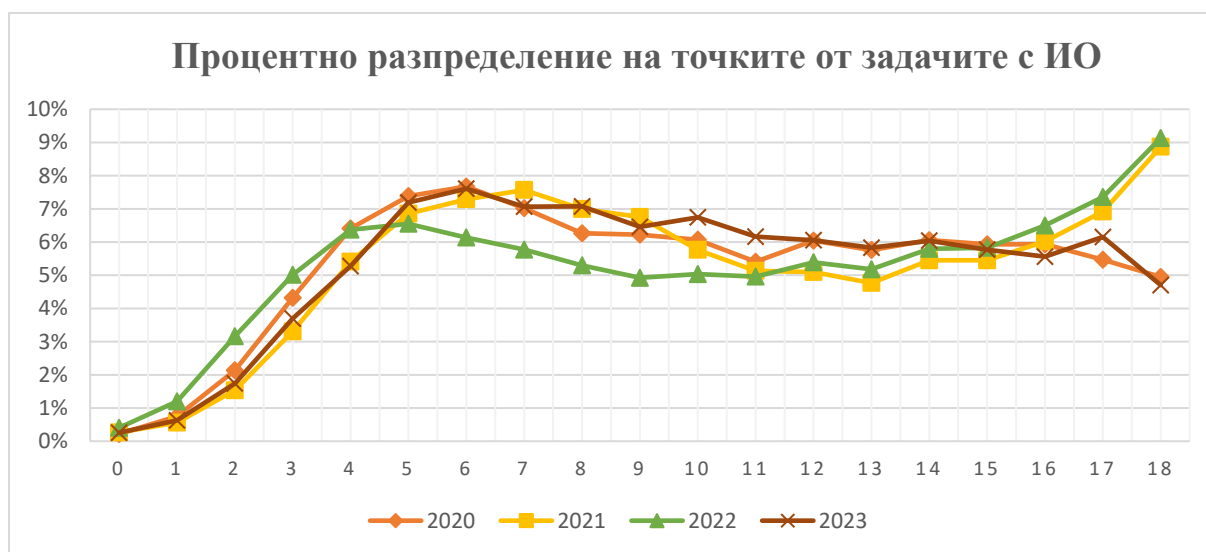
Фиг. 1.32. Процентно разпределение по области на точките от задачи от 21 до 23 с разширен СО – 2023 г.

- направен е извод, че задачите със СО не допринасят съществено за увеличаване на баловете на учениците, получени от задачите с ИО. Това важи за почти всички области, с изключение на София-град. Една от причините за това е, че в по-малките населени места конкуренцията при кандидатстване в гимназиите е

сведена до минимум, докато в София-град все още има конкуренция. Този извод се подкрепя от диаграми като тази за 2023 г. на фигура 1.32, в която по хоризонтала са дадени точките, получени от СО, а по вертикала е процента на учениците, които са получили тези точки.

Обобщения и изводи

В последната точка на тази глава са направени някои обобщения върху задачите с ИО от изследваните 4 години. От фигура 1.33 се вижда, че процента учениците, получили точки в интервала $[0;5]$ се изменя по един и същи начин през четирите години – той нараства с почти еднакви темпове при увеличаване на точките. След това, в интервала $[6;16]$ има постоянни стойности, а в края на интервала там, където са най-големите балове има различна характеристика за две двойки години. За 2020 и 2023 г. процентът намалява, докато за 2021 и 2022 г. той нараства. Не може да не се отбележи, че 2020 беше годината, в която започна **обучението от разстояние в електронна среда** (ОРЕС), а 2023 година учениците вече бяха приключили с това обучение. За средните две години – ОРЕС беше основния начин на обучение.



Фиг. 1.33. Процентно разпределение на точките от 2020 до 2023 г. на задачите с ИО

В тази глава са предложени два варианта за трансформация на точките в оценка по шестобалната система – линеен и кумулативен. И по двата варианта се наблюдават по-високи стойности на слабите оценки за 2022 г., което важи и за отличните оценки. Една от причините за това би могло да е двугодишното обучение в електронна среда, в което се оказаха учениците от този випуск. За по-слабо подготвените ученици това обучение създава проблеми, докато за отличните ученици създава възможност за по-добра подготовка. Този феномен може да се обясни с това, че при онлайн обучението

самоподготовката на учениците е на преден план, което за по-добрите от тях създава възможности за по-голям индивидуален напредък. За по-слабо подготвените ученици такава самостоятелната работа създава повече трудности. Това слабо представяне на учениците от 2022 г. се потвърждава и от високия процент (68%) на тези, които са получили между 0 и 5 точки на задачите със СО. Този процент е още по-висок (82%) за по-малките области, като Плевен и Разград.

В същата точка е направена сравнителна таблица на задачите с ИО от всичките четири години по съдържателна и познавателна област. Трябва да се отбележи, че разпределението на тези задачи по двете области е направено от автора на дисертационния труд, а не от комисията, която ги е съставила, т. к. това не е общодостъпно. Прави впечатление, че 11 от всички 27 теми от съдържателните области или не са представени в задачите с ИО, или има само 1 задача за четирите години. От друга страна, голяма част от тях присъстват в задачите със СО и може да се приеме, че Националните външни оценявания покриват цялото съдържание, което е описано в спецификациите им. Според тази таблица, почти всички задачи с ИО са от познавателно равнище Знание или Разбиране, по изключение има задача от ниво Приложение. Това е очаквано за задачи с избираем отговор.

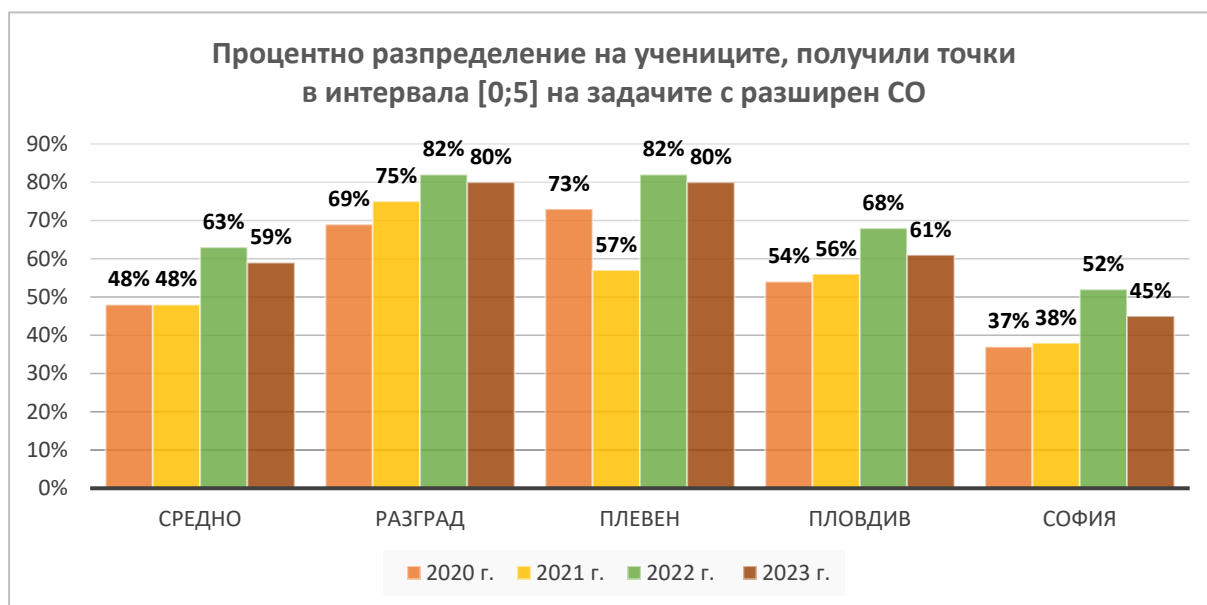
	α на Кронбах	SEM
2020 г.	0,8604	1,7476
2021 г.	0,8677	1,7466
2022 г.	0,8867	1,7184
2023 г.	0,8485	1,7762

Таблица 1.66. Коефициент на надеждност на тестовите задачи с ИО

В края на главата, според коефициента на надеждност на задачите с ИО от четирите НВО в таблица 1.66 е направен извод, че такъв тест е подходящ:

- за изследване на резултатите по математика на учениците от VII като цяло. Изводите, които са направени с висока степен на надеждност важат за цялата съвкупност;
- за сравняване на тези резултати по различни големи групи като: пол, регион, училища.

За всички години $\alpha < 0,95$ и стандартната грешка е такава, че 95 процентния доверителен интервал е с дължина 5-6 точки. Това дава основание да се каже, че задачите с ИО не са подходящи, ако трябва да се вземат решения, свързани с класиране на учениците за прием след VII клас. За тази цел, задачите със СО би трябвало да повишат тази надеждност. Да разгледаме дали това е така за учениците като цяло и в отделните области в страната.



Фиг. 1.36. Процентно разпределение по области на учениците, получили точки в интервала [0;5] на задачите с разширен СО

На фигура 1.36 е даден процентът на учениците, които са получили точки в интервала [0;5]. Средно за страната около 50% от учениците от първите две години и около 60% от следващите две години или не са получили точки, или те са минимален брой от задачите с разширен СО. Ако се разгледат тези данни по области, се вижда, че за по-малките области Разград и Плевен тези проценти са по-големи с още 20%. За област Пловдив са около средните за страната, а за област София-град са по-ниски от средните. Причина за разликата в представянето е, че за по-малките области конкуренцията за прием след VII клас е или много слаба, или изобщо я няма. Докато за по-големите градове, особено за София, тази конкуренция е много висока. Следователно за голям процент от учениците в страната задачите с разширен СО не носят добавена стойност към задачите с ИО.

Друго наблюдение, което може да се направи от фигура 1.36, е рязкото покачване на процента на учениците с ниски резултати по втори модул на НВО за 2022 г. (от 48%

на 63% средно). Като се вземе предвид, че разпределението според коефициента на трудност на задачите с ИО през четирите години не е много различно, то проблемът с показването на тези проценти може да се потърси в обучението в електронна среда или в някакви големи разлики в естеството на задачите със СО. Ясно е, че обучението от разстояние в електронна среда не е подходящо в комуникацията между учител и ученик. Това е особено важно, когато ученикът трябва да опише решение на задача, а учителят да коментира това решение, да посочи слабостите в него, да даде подпомагащи въпроси и пр. Намалването на този процент през 2023 г., когато учениците отново бяха в класните стаи потвърждава този извод. Интересно е да се проследи тази тенденция и през следващите години.

Задачите с разширен СО през изследваните години нямат съществени разлики. Те са от основните три типа:

- алгебрична задача, в която иска да се реши линейно уравнение и линейно неравенство, като се направят няколко тъждествени преобразувания.
- задача за моделиране, като през 2020 и 2022 г. е била от типа „движение“, през 2023 г. е от тип „работа“, а през 2021 г. при дадена диаграма трябва да се моделира уравнение и да се правят различни изводи от нея.
- геометрична задача, в която освен основни геометрични знания, трябва да се прилагат и допълнителни аналитични разсъждения.

ВТОРА ГЛАВА. Анализ на НВО по математика след VII клас чрез IRT

Основни принципи и модели

В началото на тази глава са описани основните принципи на теорията на тестовете, базирана на вероятностното моделиране – Item Response Theory.

Класическата теория на тестовете е широко използвана в България поради нейната прозрачност и достъпност, но има известни ограничения. Ето две от тях:

- тестовите балове зависят от избраните задачи в теста (тяхната трудност);
- определянето на трудността на задачите зависи от конкретните изпитвани;

В научната литература от десетилетия се развива нова теория, базирана на вероятностното моделиране. Тя е известна като *Item Response Theory (IRT)*. В теорията на вероятностното моделиране има разработени различни модели. Един от най-широко

използваните модели е описан от Г. Раш. Неговият труд (Rasch, 1980) излиза за пръв път през 1960 г. Този модел се нарича още *еднопараметричен логистичен модел* или *Rasch-model* и може да се разглежда като частен случай на модела на А. Бирнбаум описан през 1968 г. в (Birnbaum, 1968) и в последствие развит от Hambelton и Swaminathan в книгата (Hambelton, R., Swaminathan, H., Rogers, H., 1991). В българската литература е описан от няколко автори, например от Ф. Стоянова в книгата (Бижков, 1996), (Банков, 2002), (Джалев, 2014) и (Джалев, 2013). Използването на вероятностни модели е популярно в национални изследвания на някои държави и особено в международни такива (TIMSS, PISA, PIRLS и др.). В последните години има публикации и в България – (Банков, 2012), (Алашка, 2016) и (Алашка, 2017), (Цонев, 2023-а), (Tsonev, 2024) и (Цонев, 2024-а).

Някои от предимствата на IRT са:

- мярката на измерването на параметрите (особено в модела на Раш) дава възможност за анализиране на качествените данни с помощта на количествени методи;
- оценката на трудността на задачите не зависи от избора на изпитваните, от които тя е получена;
- оценката на нивото на способностите на изпитваните не зависи от използвания набор задачи, с които те са изпитани;
- наличието на непълнота на данните не е критичен недостатък при изследването на тестовете.

Три от основните предположения в IRT са (Ким, 2007):

- 1) съществуват скрити (наречени още *латентни*) параметри на личността, които са недостъпни за непосредствено наблюдение;
- 2) съществуват достъпни (наречени *индикатори*) за наблюдение параметри, които са свързани с латентните параметри. По значението на тези индикатори може да се съди за значението на латентните параметри;
- 3) оценяването на латентния параметър трябва да бъде едномерно. В частност, това означава, че тестът трябва да измерва знанията само в една ясно зададена предметна област. Ако условието за едномерност не е изпълнено, то е необходимо

тестът да се преработи, като се отделят задачите, които нарушават неговата хомогенност.

Въпросът за едномерността на 18-те тестови задачи с ИО от Националните външни оценявания, които се разглеждат в настоящия дисертационен труд, е демонстриран върху проведеното през 2020 г. НВО по математика. Използван е факторен анализ и софтуерът SPSS. Оказва се, че има два компонента в теста, от които първият е съществен „силен“, т.е. тестът е предимно едномерен (оценява един конструкт), но има три задачи, които оценяват нещо по-странично от този конструкт, който да може да се нарече „втори конструкт“ (втора размерност). Най-голям принос към този втори конструкт имат задачите с номера 7, 15 и 18. Те са разгледани по-подробно в дисертационния труд и са намерени причини, които отделят тези задачи от общото поведение на теста.

За разлика от приетото в България точкуване на тестовите задачи, за да бъдат те изследвани с IRT, трябва да се точкуват според вида на задачите така:

- при задачите с ИО – с 1 точка за вярно посочен отговор или с 0 точки при грешен, непосочен отговор или при посочени повече от един отговор;
- при задачите със СО – с цяло число точки в интервала $[0, n]$. Това число n е равно на броя *важни стъпки* в решението на задачата. Стойността на n се определя от авторите на задачата и показва колко са тези важни стъпки в нейното решаване. Тази идея е описана от (Банков, 2023). При това, ако задачата е тип Вярно-Невярно, то $n = 1$. Най-често $n = 2$ или $n = 3$.

За краткост, по-надолу задача, при която $n = 1$ ще се нарича *0-1 задача*, ако $n = 2$ ще нарича *0-1-2 задача* и така нататък.

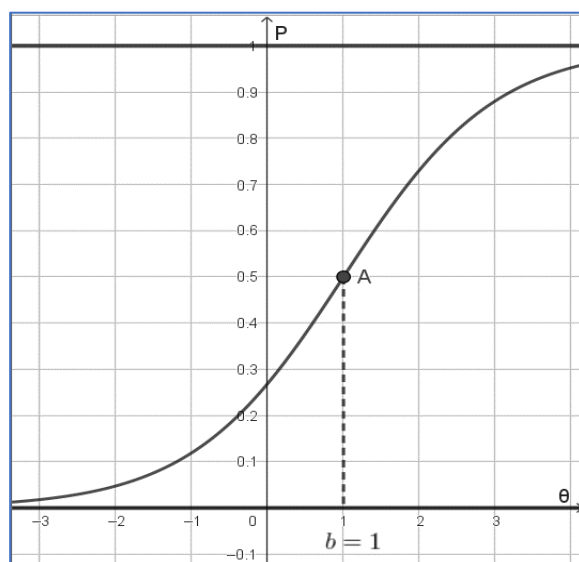
За да се обясни модела на Раш, ще се разгледат задачи от вида 0-1. Основната идея на модела е, че възможността един ученик да реши дадена задача може да се измери чрез абстрактна величина, наречена *способност*, която се означава с θ . От друга страна на всяка задача може да се съпостави така наречената *характеристична функция*, чиято графика се нарича *характеристична крива*. Тази функция описва вероятността задачата да бъде правилно решена в зависимост от същата абстрактна величина θ , т.е. тази функция изразява вероятността ученик с дадена способност да реши правилно задачата. Стойността на θ , при която вероятността задачата да бъде правилно решена е $1/2$ се нарича *трудност* на задачата и се означава с b .

Предположението, което се прави в Раш-модела е, че способността θ на ученика да реши задачата и трудността b на тази задача могат да се разположат на една и съща скала и да се измерват с едни и същи единици, наречени *логити*. Тази скала се нарича още *скала на способностите (Ability scale)*. Успехът на ученика е функция на разликата $\theta - b$. Ако тази разлика е положителна и голяма, това означава, че има голяма вероятност ученикът да реши съответната задача вярно. Ако тази разлика е отрицателна и голяма по модул, то има голяма вероятност ученикът да не реши вярно дадената задача.

Ако задачата е 0-1, характеристикната ѝ функция има вида

$$P(\theta) = \frac{e^{\theta-b}}{1 + e^{\theta-b}},$$

където b е параметър, който се нарича *трудност* на задачата.



Фиг. 2.1. Характеристична крива на Rasch model на задача 0-1

Графиката на такава крива при $b = 1$ е изобразена на фигура 2.1. Координатите на точката от характеристикната крива с ордината $P = 0,5$ (в случая точка A) показват, че ученик със съответната способност (в случая $\theta = 1$) има вероятност 50% да получи верен отговор на тази задача. Тази стойност на θ (в случая $b = 1$) е трудността на задачата. Обикновено трудността на задачите е в интервала $(-2; 2)$. Ако b е близо до -2 , значи задачата е лесна, а ако е близо до 2, значи задачата е трудна.

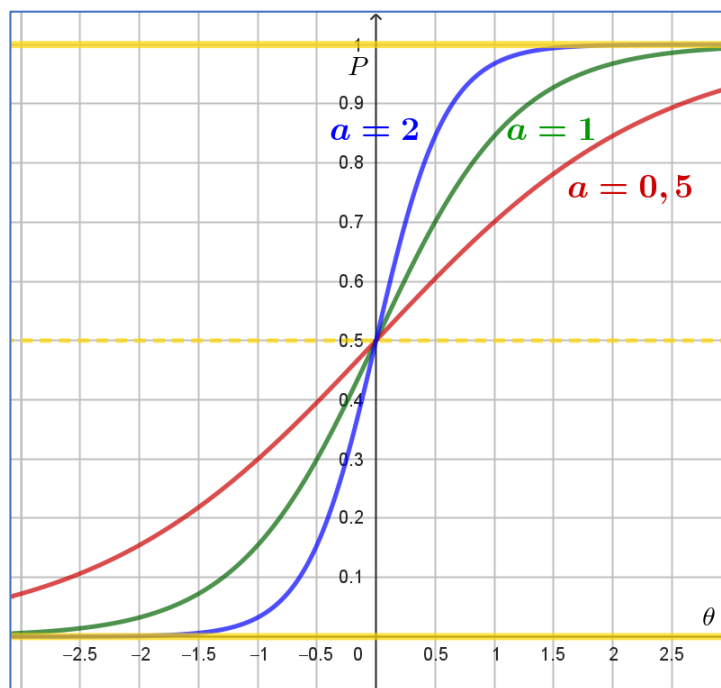
Моделът, предложен от Раш, в последствие е развит и през 1968 г. Бирнбаум (Birnbbaum, 1968) въвежда *двупараметричен* и *трипараметричен* модел. При първия

модел, освен параметъра b , се въвежда нов параметър a . Характеристичната функция на всяка задача има вида:

$$P(\theta) = \frac{e^{1,7a(\theta-b)}}{1 + e^{1,7a(\theta-b)}},$$

където параметърът a има смисъл на *дискриминация* на задачата. Той задава наклона на съответната характеристична крива в точката на трудност b на задачата. Колкото тази крива е „по-стръмна“, толкова съответната задача по-добре разграничава силните от слабите ученици, които имат способности близки до b . Най-често интервалът, в който приема стойности a е $(0;2)$. Този модел ще бъде означен като *2PL-модел*. В известен смисъл той е обобщение на еднопараметричния Раш-модел, който ще се нарича още *1PL-модел*.

На фигура 2.4 са изобразени характеристичните криви на задачи с трудност $b = 0$, които имат дискриминация $a = 0,5$, $a = 1$ и $a = 2$.



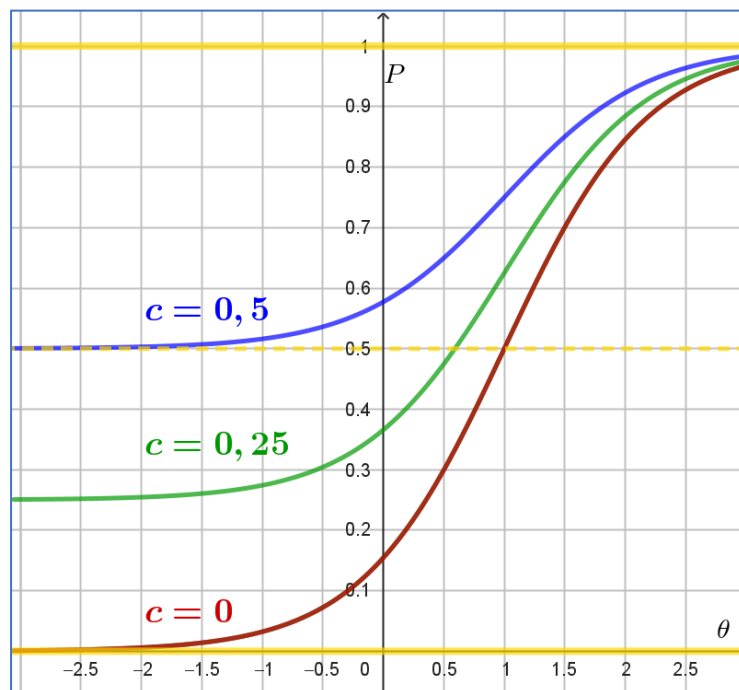
Фиг. 2.4. Примери за 2PL модел с три варианта на параметъра a

При трипараметричния модел (*3PL-модел*) се въвежда и параметър c , който се счита за коефициент на налучкване. Характеристичната функция на този модел е:

$$P(\theta) = c + (1 - c) \frac{e^{1,7a(\theta-b)}}{1 + e^{1,7a(\theta-b)}}.$$

Графично стойността на параметъра c изобразява хоризонтална асимптота $P = c$ на характеристичната крива при $\theta \rightarrow -\infty$.

На фигура 2.5 са изобразени графиките на характеристични криви на три задачи с трудност $b = 1$, дискриминация $a = 1$ и различни параметри на налучкване $c = 0, c = 0,25$ и $c = 0,5$.



Фиг. 2.5. Примери за 3PL модел с три варианта на параметъра c

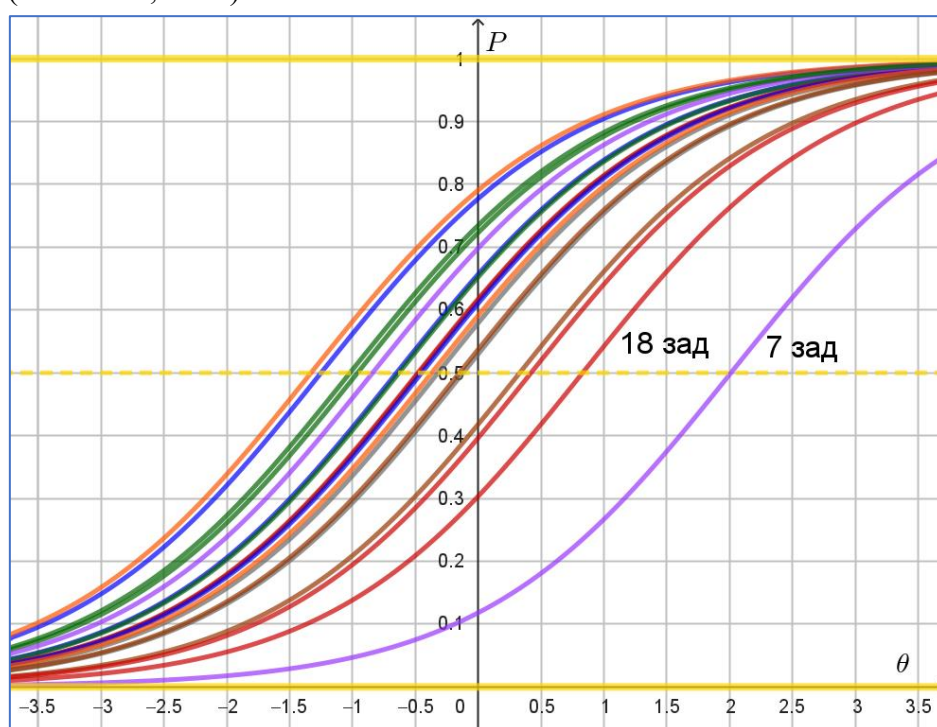
Избор на подходящ модел за изследване на задачите с избираем отговор

Използването на Раш-модела за изследване на даден тест предполага, че задачите, включени в него, са с еднаква дискриминация и вероятността за налучкване на верния отговор е минимална. Чрез този модел може да се установи кои задачи не се съгласуват достатъчно добре с емпиричните данни, които се получават след като се проведе теста. При него постановката е, че не моделът трябва да съответства на емпиричните данни, а данните трябва да съответстват на модела (Ким, 2007). Обикновено се препоръчва проблемните задачи да не се включват в следващи тестове, преди да се преработят.

При моделите с 2 и 3 параметъра подходът е друг. Чрез тях се търси модел, който най-добре да съответства на изследвания тест.

Във втора точка на тази глава за всяка задача с ИО и за всяко НВО от изследваните четири години чрез приложението jMetrik (Meyer, 2024) са анализирани получените характеристичните функции според трите модела на IRT.

В първата подточка на тази точка са разгледани характеристиките на задачите с ИО от НВО, проведено през 2020 г. За настоящото изследване са използвани резултатите от проведеното НВО в края на VII клас по математика с 18 836 ученици от изследваните четири административни области на България, които са се явили на изпит през 2020 година. Нека се отбележи отново, че тези области са София-град, Пловдив, Плевен и Разград. Обработени са данните за задачите с избираем отговор, които са първите 18 в този тест. Използваният софтуер jMetrik оразмерява скалата на способностите (Ability scale) така, че разпределението на способностите на всички ученици е със средна стойност 0 и стандартно отклонение 1. Графиките са начертани с приложението GeoGebra (GeoGebra, 2024).



Фиг. 2.6. Характеристични криви на задачите от 2020 г. – 1PL

На фигура 2.6 са дадени характеристичните криви на всичките 18 задачи с ИО според еднопараметричния Раш-модел (1PL). От нея се вижда, че тези задачи покриват диапазона $(-3,5; 3,5)$ на нивото на способностите θ . Кривите на някои задачи се припокриват, например 1 и 3; 9, 11 и 12; 15 и 16. В тази връзка някои от тях, например 3, 11, 12 и 16, могат да бъдат премахнати от теста, без да се нарушат съществено измерителните му качества. В същото време се наблюдават липсващи задачи в някои интервали на трудност (гледаме нивото $P = 0,5$). Чрез допълнителна преработка би могло да се потърсят задачи с трудност в интервалите $(-0,12; 0,33)$, $(0,42; 0,83)$ и $(0,83; 2,01)$. Очевидно най-трудни са задачи 7 и 18, което се потвърждава и от

Класическата теория на тестовете. Тези задачи са разгледани подробно в първа глава на дисертационния труд.

В таблица 2.3 са дадени стойностите на параметрите според трите модела на параметризиране. Впечатление правят две задачи с номера 15 и 9, които имат високи стойности на налучкване, съответно $c = 0,37$ и $c = 0,29$. Тези задачи не се отличават по никой от показателите на Класическата теория на тестовете и са разгледани в тази глава.

	1PL	2PL		3PL		
№	b-param	a-param	b-param	a-param	b-param	c-param
1	-0.65	2.66	-0.44	2.96	-0.23	0.10
2	0.33	1.24	0.25	2.38	0.68	0.19
3	-0.63	2.01	-0.45	2.58	-0.13	0.16
4	0.42	1.12	0.35	2.72	0.81	0.22
5	-1.25	2.29	-0.80	2.25	-0.72	0.04
6	-1.33	1.89	-0.89	2.03	-0.68	0.14
7	2.01	1.12	1.74	2.94	1.51	0.07
8	-0.32	1.21	-0.29	1.98	0.33	0.26
9	-0.47	1.44	-0.39	2.73	0.25	0.29
10	-0.96	2.10	-0.64	2.93	-0.22	0.24
11	-0.48	1.80	-0.36	2.98	0.12	0.24
12	-0.45	2.07	-0.34	2.41	-0.12	0.10
13	-0.37	1.77	-0.29	2.93	0.15	0.22
14	-0.84	1.97	-0.58	2.21	-0.35	0.12
15	-0.12	0.80	-0.13	2.97	0.85	0.37
16	-0.15	1.70	-0.15	2.66	0.24	0.18
17	-1.02	1.27	-0.84	1.60	-0.29	0.26
18	0.83	0.57	1.22	2.57	1.48	0.25

Таблица 2.3. Стойности на параметрите според трите модела на задачите от 2020 г.

В следващият етап на анализа на теста се проверява доколко трите модела на параметризация (1PL, 2PL и 3PL) съответстват на емпиричните данни. Ще приложим графичен метод, който е описан на стр. 66-67 в книгата (Hambelton, R., Swaminathan, H., Rogers, H., 1991). За всеки от трите модела всеки от учениците получава стойност на параметъра θ (ability), който е число в интервала $(-3;3)$. Този интервал се разделя на 12 равни части. Ако средата на всеки подинтервал означим с θ_i ($i = 1, 2, \dots, 12$), то на всички ученици със способност θ в този подинтервал приемаме, че имат средната стойност θ_i . След това за всяка задача и за всеки подинтервал се намира отношението на броя на учениците r_i , които са отговорили вярно на задачата към броя на всички ученици m_i в този подинтервал. Така се получава величината

$$p(\theta_i) = \frac{r_i}{m_i},$$

която се явява експерименталната стойност на вероятността да се даде верен отговор на тази задача на ученик със способност θ_i .

1. Стойността на израза $200 - 20 \cdot \left(-2 \frac{1}{2}\right)$ е:

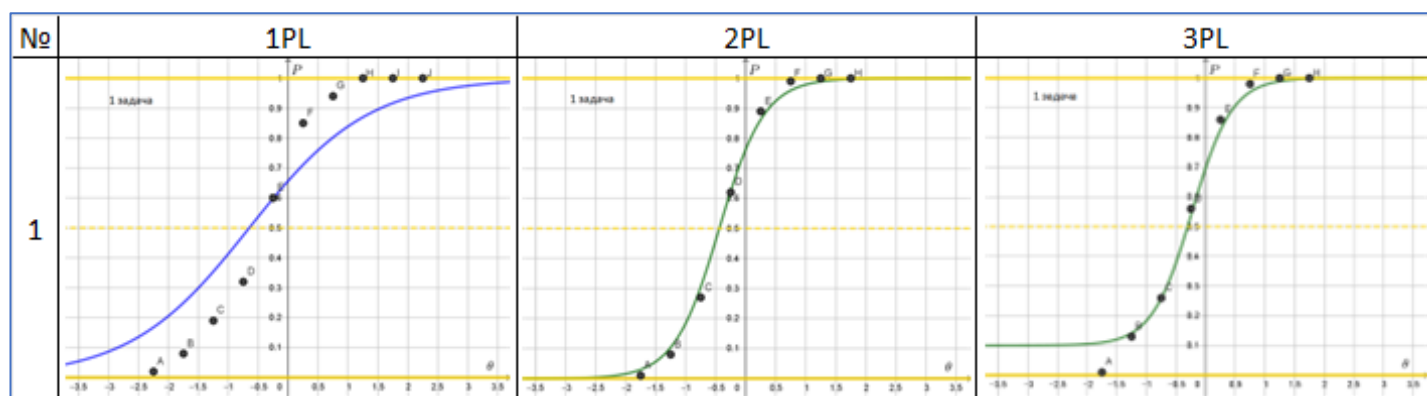
А) –450

Б) –72

В) 208

Г) 250

Фиг. 2.7. Условие на задача 1 от 2020 г.



Фиг. 2.8. Графика с остатъците и характеристична крива на задача 1 от 2020 г.

Чрез следващата задача ще демонстрираме идеята с която се изследва съответствието на теоретичните криви и емпиричните данни. На фигура 2.7 е дадено условието на задача 1. Под надписите 1PL, 2PL и 3PL на фигура 2.8 са разположени характеристичните криви на тази задача според трите модела. Емпиричните данни за всеки модел са представени с точките А, В, С, ..., които имат координати съответно $A(\theta_1, p(\theta_1))$, $B(\theta_2, p(\theta_2))$, $C(\theta_3, p(\theta_3))$, ... Тези графики се наричат още *графики с остатъците*. Даден модел има добро съвпадение с данните, ако графиките с остатъци са близки до съответните характеристични криви на задачите.

За тази задача може да се каже, че:

- тя е с оптимална трудност – графиките пресичат хоризонталната права $P = 0,5$ в точка с абсциса около $-0,5$;
- графиките на дву- и три-параметричния модел са по-изправени, т. е. тя има по-добра разграничителна сила. За двата модела стойността на параметъра на дискриминация е $a = 2,66$ (за 2PL) и $a = 2,96$ (за 3PL);

- липсват точки, съответстващи на ученици със способности по-малки от $-2,5$ и по-големи от $2,5$ за 1PL модела, а за другите два модела интервалът на способностите е още по-стеснен $(-2; 2)$;
- 1PL моделът има големи разлики между теоретичните и емпиричните резултати;
- 3PL моделът дава висок коефициент на налучкване, което не се съгласува добре при учениците с ниски способности (точка А е далеч от кривата);
- 2PL моделът се съгласува в най-голяма степен с емпиричните данни;
- задачата е от съдържателна област „Действия с рационални числа“ и такива задачи присъстват във всички НВО след VII клас. Подобна на тази задача е задача 1 от 2021 г. на фигура 1.19, като анализа на двете задачи от гледна точка на КТТ е идентичен.

По тази схема са разгледани още четири задачи с номера 7, 9, 15 и 18.

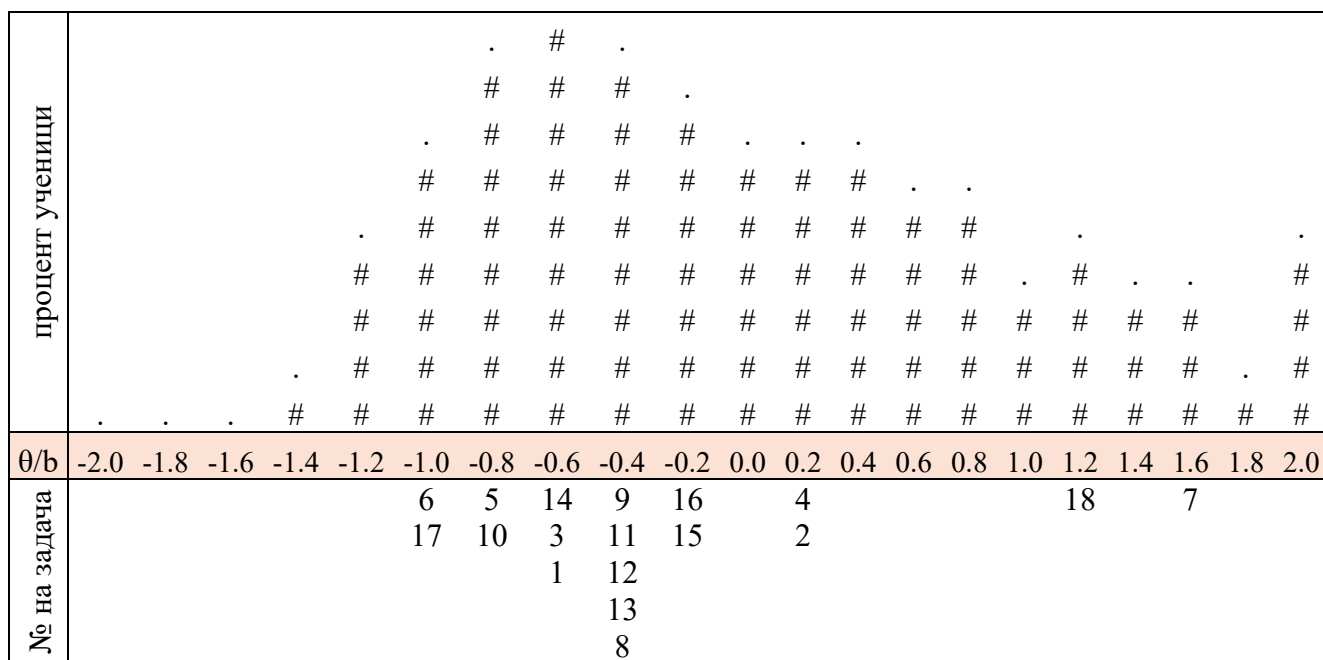
По-нататък са сравнени коефициентите на трудност и разграничителна сила на задачите с ИО, получени според двете теории. Установено е, че коефициентът на корелация е достатъчно висок в двата случая $0,98981$ за коефициента на трудност и $0,919809$ за коефициента на дискриминация.

Накрая на изследванията в тази подточка се построява така наречената *item-person-map* диаграма. За тази цел се използва, че трудността на задачите (b -параметър) и способностите на учениците (θ) могат да се разположат върху една и съща скала, наречена *Ability scale*. Отново използваме резултатите от 2PL-модела. На фигура 2.15 маркираният ред, означен с θ/b , е интервалът $(-2; 2)$ от скалата на способностите. Интервалът е разделен на подинтервали с дължина $0,2$. Под този ред са написани номерата на 18-те задачи, като всяка задача е поставена на мястото, където попада нейният параметър на трудност b . Над маркирания ред е представен процентът ученици със способност θ в съответния интервал с дължина $0,2$, като:

- на всеки символ “#” отговаря 1% ученици;
- на всеки символ “.” отговаря под 1% ученици.

Така, под маркирания ред е графиката на разпределението на трудността на задачите върху скалата на способностите, а над него е разпределението на способностите на учениците върху същата скала. Може да считаме, че тестът, съставен от представените

задачи, е добре балансиран за съответните ученици, ако двете разпределения имат връх в близки стойности на маркирания ред и покриват почти едни и същ интервал от него. Графики като фигура 2.15 дават възможност да се прецени до колко тестът, съставен само от 18-те задачи с избираем отговор, е „подходящ“ за изследваните ученици от четирите области.



Фиг. 2.15. Диаграма на задачите и респондентите от НВО – 2020 г.

Разглеждайки фигура 2.15 може да се направят следните наблюдения:

- учениците се разпределят по целия интервал от $(-2; 2)$;
- задачите са съсредоточени в средната третина на интервала. Това означава, че липсват задачи за много слабите и много силните ученици;
- и задачите, и учениците имат сравнително добро „нормално“ разпределение;
- най-много задачи имат коефициент на трудност около $-0,4$, докато най-големият брой ученици са с малко по-ниски способности, т.е. пиковите стойности са близки;
- както се показва от еднопараметричния модел и от КТТ по-леките задачи са с номера 5, 6 и 17,
- най-трудните са с номера 18 и 7 и те са разположени сравнително далеч от основната група задачи.

По тази процедура в останалите три подточки е направен анализ още 10 задачи от следващите три изследвани години, като от съответните графики са направени изводи:

- относно хомогенността на задачите във всеки тест;
- чрез сравняване на емпиричните и теоретичните данни в трите модела на параметризация на резултатите от четирите изследвани години е направено заключение, че двупараметричният модел има известно предимство;
- по отношение на коефициентите на трудност и дискриминация на задачите според КТТ и двупараметричният модел на IRT, като е установена висока степен на съгласуваност на двете теории;
- чрез *item-person-map* диаграма е показано, че трудността на задачите добре се съгласува със способностите на учениците.

Установено е, че при по-голяма част от задачите с ИО от разглежданите четири НВО има съответствие между коефициентите на трудност и на дискриминация според двете теории – КТТ и IRT, но има задачи, които показват известни различия. Освен това, с вероятностното моделиране са открити задачи с висок коефициент на налучкване. При тях са потърсени и намерени причини, които спомагат да се уцели верния отговор, без да се правят необходимите разсъждения или изчисления, които водят до него. Така се потвърждава основната хипотеза в дисертационния труд, че съчетаването на двете теории при анализ на тестовите резултати дава по-обективна, по-точна и надеждна картина.

Сравняване на резултатите на задачите с ИО за периода 2020-2023 г.

В третата точка на тази глава са сравнени резултатите на всички задачи с ИО от изследвания период 2020-2023 година. Методиката на това сравнение, описана в първата подточка, се базира на идеята да се използват няколко общи задачи (*anchor*-задачи), давани в два различни теста на две различни групи ученици и приравняване на резултатите от едната група към тези от другата. Поради липсата на общи задачи в годините, за такива са потърсени и намерени двойки задачи с подобни характеристики по отношение на трудност и дискриминация.

Във втората и третата подточка е приложена процедурата за пресмятане на резултатите от 2020 и 2022 години върху тези от 2021 г. Като извод от тази точка е установена тенденцията за понижаване на постиженията на учениците през целия изследван период. Факт, който се разминава донякъде със официалната статистика.

Причина за това е, че при нея се сравняват само средните стойности на резултатите от НВО по математика.

Приложената методика освен за това сравняване, може да бъде използвана и за създаване на банки от задачи с избираем отговор, които са „оразмерени“ върху единна скала и са групирани по коефициент на трудност, който се получава обективно, а не според виждането на авторите на тестовите задачи. Често авторът има една представа за трудността, докато обективната реалност сочи друго.

Точкуване на задачите със СО

В четвъртата точка на тази глава са разгледани задачите със СО. Точкуването на задачите със свободен отговор често е свързано с определени трудности. Две от тях са:

- Как да се определи максималният брой точки за дадена задача?
- Как да се разпределят тези точки върху решението на задачата?

В стремежа си оценяването да е максимално обективно често се дават точки, които освен цели числа, са и кратни на 0,5, дори и на 0,25.

Дали това е оправдано?

В настоящото изследване се дава една гледна точка на този въпрос, която се основава на методите на вероятностното моделиране. Разгледани са задачите, давани през 2023 г., както и получените резултати на всички 21 060 седмокласници от четирите административни области София-град, Пловдив, Плевен и Разград. Използван е доклада на автора по тази тема на 53-та Пролетна конференция на Съюза на математиците в България (Цонев, 2024-а).

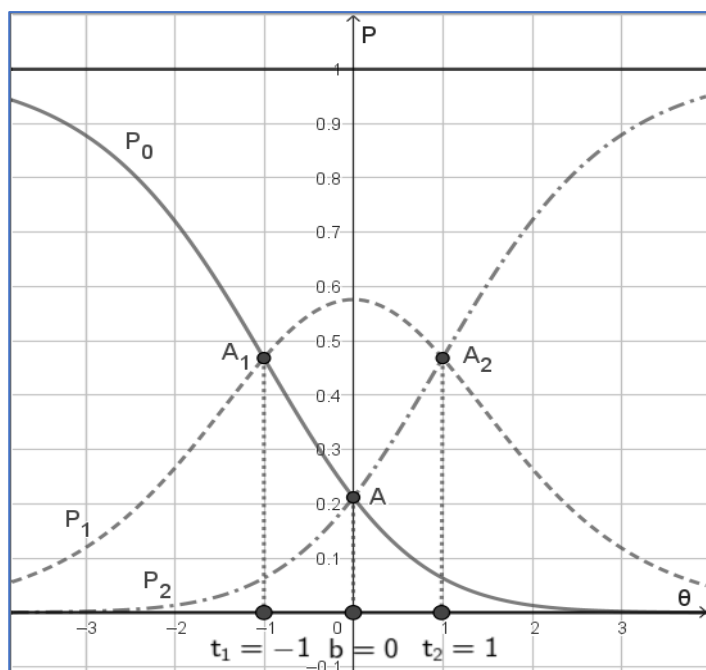
Според модела на Раш, ако една задача се точкува с 0, 1 и 2 точки, се чертаят три характеристични криви, които се описват чрез следните три функции:

$$P(X = 0|\theta) = \frac{1}{1 + e^{\theta-b-t_1} + e^{2\theta-2b-t_1-t_2}}$$

$$P(X = 1|\theta) = \frac{e^{\theta-b-t_1}}{1 + e^{\theta-b-t_1} + e^{2\theta-2b-t_1-t_2}}$$

$$P(X = 2|\theta) = \frac{e^{2\theta-2b-t_1-t_2}}{1 + e^{\theta-b-t_1} + e^{2\theta-2b-t_1-t_2}}$$

Първата от тях показва вероятността за получаване на грешен отговор ($X = 0$), втората показва вероятността за получаване на частично верен отговор ($X = 1$), а третата – вероятността за получаване на правилен отговор ($X = 2$).

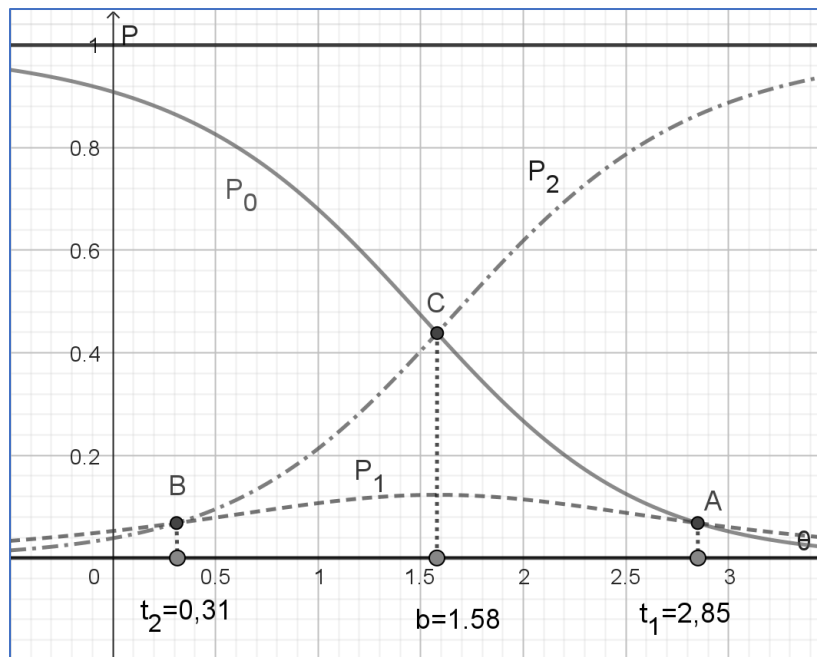


Фиг. 2.2. Характеристични криви на Rasch model на задача 0-1-2

На фигура 2.2 е даден пример, в който графиките P_0 , P_1 и P_2 съответстват на вероятността за получаване на 0, 1 и 2 точки. Трудността на такава задача се получава от абсцисата на пресечената точка A на P_0 и P_2 (в случая $b = 0$). Интерес представляват абсцисите t_1 и t_2 на пресечните точки A_1 и A_2 съответно на P_0 , P_1 и на P_1 , P_2 (в случая $t_1 = -1$ и $t_2 = 1$). Изводът, който може да се направи при такава задача е, че учениците със способности по-малки от $t_1 = -1$ е най-вероятно да получат 0 точки, тези със способности между $t_1 = -1$ и $t_2 = 1$ е най-вероятно да получат 1 точка, а тези със способности над $t_2 = 1$ е най-вероятно да получат 2 точки.

Проблемни задачи са такива, при които $t_1 > t_2$. Такъв пример е даден на фигура 2.3. Както се вижда графиката P_1 се намира „под“ графиките на P_0 и P_2 . Това означава, че няма интервал от способностите, в който вероятността ученикът да получи 1 точка да е най-голяма. Така, че такава задача е по-подходящо да е от вида 0-1.

По подобен начин се описват и задачи с повече междинни точки, като функциите от тип „камбана“ са повече.



Фиг. 2.3. Характеристични криви на Rasch model на задача 0-1-2 при $t_1 > t_2$

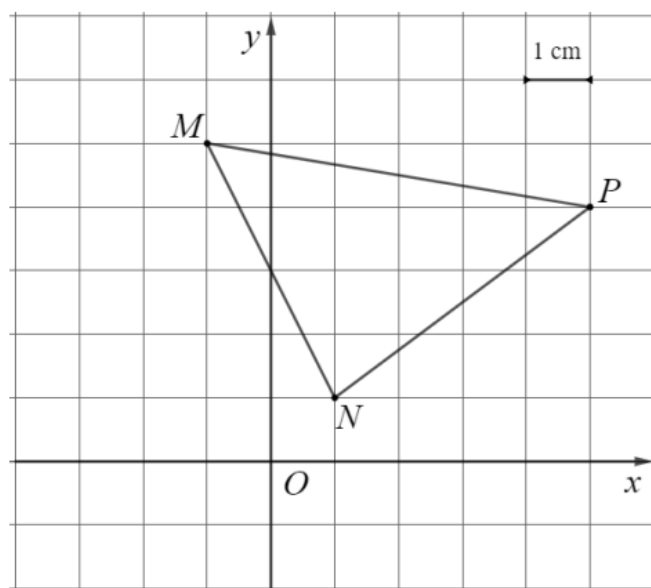
За да се приложи вероятностният метод, точкуването на задачите със свободен отговор от 2023 г., което е използвано от оценителите на НВО, трябва да се промени. Могат да се отделят два типа такива промени.

Първият тип е при задачите с номера 19 А), 19 Б) и 20 В), които са оценявани съответно с 4, 3 и 4 точки за верен отговор и 0 точки за грешен или непосочен отговор, точкуването се променя на 1 точка за верен отговор и 0 в другите случаи. Така те стават така наречените 0-1 задачи.

Вторият тип е при останалите задачи, в които точкуването е много раздробено. Такова оценяване не е приложимо за изследване с методите на IRT. По тази причина е необходимо броят на точките в тяхното оценяване да е цяло число и да е сравнително малко. В (Банков, 2023) е анализиран начинът на точкуване на задачите, които оценяват частичния напредък на ученика при решаването им. Въвежда се понятието важна стъпка. Целта на важните стъпки е да се разграничат смислените моменти при решаването на дадена задача. Тези моменти се определят от автора на задачата.

Пример за задача с разширен свободен отговор, която е разделена на много важни стъпки е задача 20 А), условието на която е на фигура 2.52.

20. В координатна система Oxy с единична отсечка 1 cm са дадени точките M , N и P .



А) Запишете координатите на точките M , N и P .

Б) Запишете координатите на точка Q , симетрична на точката N спрямо координатното начало.

Фиг. 2.52. Условие на задача 20 А) и Б) от 2023 г.

Точки	0	0,25	0,5	0,75	1	1,25	1,5	1,75	2	2,25	2,5	2,75	3
%	34,7%	0%	1,6%	0%	3,9%	0%	2,6%	0%	4,1%	0%	6,4%	0%	46,6%

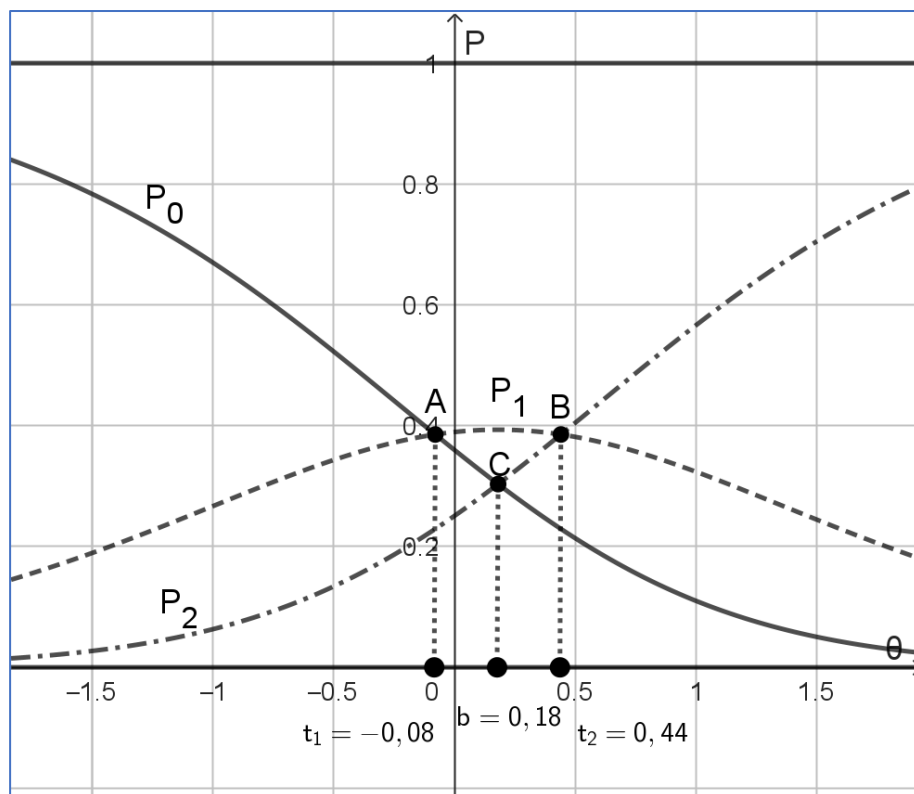
Таблица 2.19. Оценяване на задача 20 А) според МОН

Оценяването на тази задача е през 0,25 единици. В таблица 2.19 са дадени точките и процента на учениците получили тези точки. Както се вижда, процентът на учениците получили някои точки е близко до 0%. Това дава основание да се отделят 3 важни стъпки, които позволяват да се направи промяна в точкуването. Тези стъпки са:

- получени са 1 или 2 верни координати
- получени са 3 или 4 верни координати
- получени са 5 или 6 верни координати

За всяка изпълнена стъпка се получава по 1 точка и по този начин задачата става 0-1-2-3 и може да бъде изследвана с IRT.

При така приетите точкувания и стандартните точкувания 0-1 на задачите с ИО е приложен Раш-моделът на всички задачи от 2023 г.

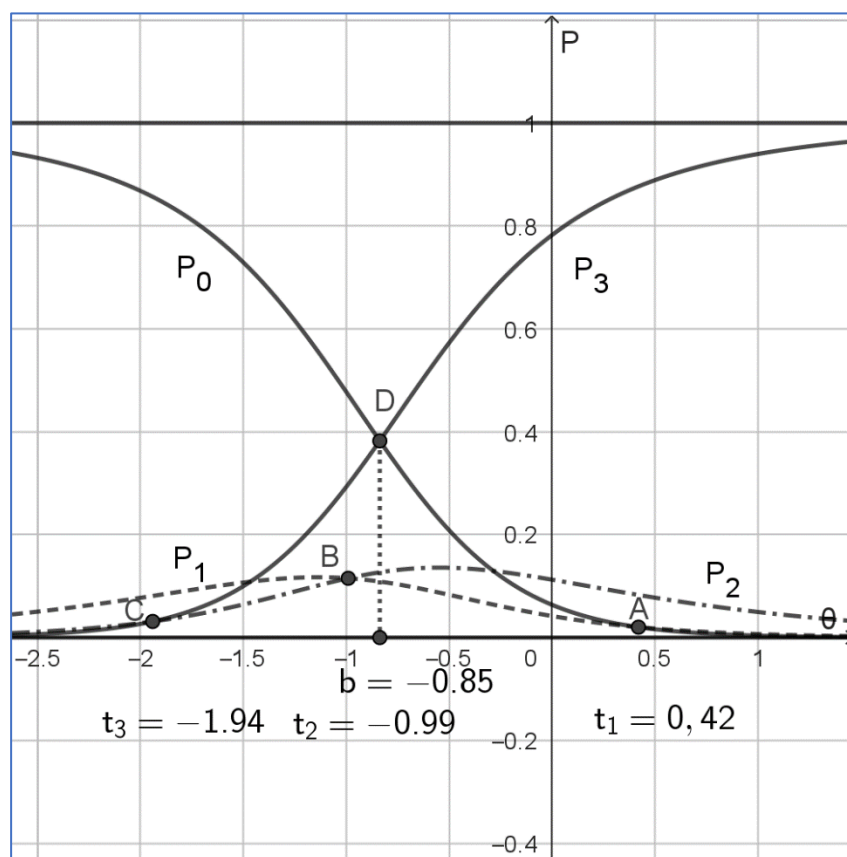


Фиг. 2.53. Характеристични криви на задача 20 Б)

При някои от задачите се получава добро разпределение на характеристичните криви (това се получава при $t_1 < t_2$). Такава е задача 20 Б), чиито характеристични криви са показани на фигура 2.53. За нея може да се каже, че учениците със способности по-малки от $-0,08$ е най-вероятно да получат 0 точки, тези със способности между $-0,08$ и $0,44$ е най-вероятно да получат 1 точка, а тези със способности над $0,44$ е най-вероятно да получат 2 точки. Тази задача има оптимална трудност ($b = 0,18$) и е с добри характеристики по отношение на разслояването на учениците по трите точки на оценяването.

При разгледаната по-горе задача 20 А) се оказва, че дори окрупненото точкуване е излишно. Според характеристичните ѝ криви на фигура 2.56 може да се каже, че учениците е най-вероятно да имат 0 или 3 точки, а междинните точки са малко вероятни. С други думи, ако един ученик може да намери координатите на една точка, той може да намери координатите на трите точки от условието на задачата. Тази задача е пример за ситуация, при която точките на пресичане на кривите P_0 , P_1 , P_2 и P_3 са подредени в обратен ред, т. е. $t_1 > t_2 > t_3$. Изводът, който може да се направи е, че тя би трябвало да се точкува с 0-1 точки. Освен това, трудността на задачата е $b = -0,85$, което означава,

че е от лесните задачи в теста и раздробяването при точкуването ѝ на 13 части (според указанията на МОН в таблица 2.19) е неоправдано.



Фиг. 2.56. Характеристични криви на задача 20 A)

По подобен начин са изследвани всички задачи със свободен отговор. Като са взети предвид графиките на характеристичните им криви е коригиран броят на важните стъпки при решаването им. Оказва се, че се открояват най-много 2 важни стъпки. На тази база може да се приложи примерно ръководство за точкуване, показано в таблица 2.21.

Тази идея за оценяване на задачите със свободен отговор би могла да бъде приложена при бъдещи НВО. Така проверителите ще дават цяло число точки за всяка важна стъпка (не повече от 2-3 на задача). Тези стъпки се определят предварително от авторите на теста. След това, екип от експерти ще направи анализ на получените резултати и според получените коефициенти на трудност ще бъдат дадени коефициенти за всяка задача, за да се получи необходимата сума от краен брой точки (в случая 100). Съвременните технологии позволяват обработката на данните да става в рамките на няколко часа. Работата на експертите също не би трябвало да отнеме повече от 1 работен ден.

В следващата глава е описан експеримент, в който се прави оценяване по традиционния начин и по начина, описан в тази точка на дисертационния труд на примерна тема за НВО след VII клас по математика.

№	Изпълнена дейност	Предложение
19А	получено 3	1
19Б	получено 10	1
20А	получени верни координати на поне 1 точка	1
20Б	- получена 1 вярна координата на Q - получена е втора вярна координата на Q	1 1
20В	получено $S = 11$	1
21А	намерено разлагане и решено полученото уравнение	1
21Б	получени корени 3 и 9	1
21В	- сгрешено неравенство, но верни изводи за получените корени в получения от неравенството интервал - верни изводи, че 3, 6 и 9 са решения на неравенството	1 1
22А	- въведени неизвестни x и $x - 2$ - получени брой работници 8 и 6	1 1
22Б	намерени 5 и 4 дни	1
23А	- правдоподобен чертеж и намерен ъгъл 40° или 100° - намерени и трите ъгъла 40°, 40° и 100°	1 1
23Б	- установено едно равенство, с което се доказва ромб - завършено е доказателството на ромб	1 1
23В	доказано $AL = BQ$	1
23Г	доказано $AP > PQ$	1

Таблица 2.21. Примерно точкуване на задачите със СО от 2023 г.

ТРЕТА ГЛАВА. Експеримент за апробация на тест за НВО по математика след VII клас

Постановка на задачата

В първата точка се представя концептуалния модел на **педагогическия експеримент**, който се проведе с 487 ученици в различни видове училища от 9 областни града на страната, на които предстоеше са завършат VII клас през 2024 г. Във връзка с възникналите въпроси в предните две глави е формулирана следната **цел**: да се приложат и сравнят методите на КТТ и IRT за анализ на тестови задачи за формиране на тестовия бал на учениците при НВО по математика след VII клас. За постигането на тази цел са поставени **изследователски задачи**:

- да се подготви изследователски инструментариум, включващ подходящи измерителни инструменти;
- да се проучат възможностите за сътрудничество с учители в страната;
- да се създаде организация за събиране на необходимите ресурси, които ще се използват в реализацията на експеримента;
- да се събере, анализира и обобщи емпиричния материал, по който да се направят съответните изводи относно методите за балообразуването на учениците.

Формулирана е **хипотеза** на експеримента:

Използването на вероятностното моделиране предлага по-прецизно изследване на качествата на тестовите задачи и спомага за въвеждане на по-опростено и по-обективно оценяване на учениците при НВО по математика след VII клас.

Този експеримент премина през три основни етапа – **предварителен, основен и заключителен**.

В **предварителния** етап са подбрани средно по 5 задачи с ИО от проведените четири НВО в периода 2020-2023 години. Критериите, по които са избирани тези задачи са:

- да са от различни съдържателни области;
- да обхващат по възможност по-голяма част от учебното съдържание V-VII клас;
- да са с различен коефициент на трудност – лесни, оптимални и трудни;
- да са показали сравнително лоши характеристики при изследването им с КТТ и IRT.

При задачите със СО, поради настъпилите промени във вида на НВО, са разглеждани само тези с разширен СО. За тях е проучено:

- какви типове такива задачи са давани през изминалите НВО;
- как са формулирани – с няколко подточки, без чертежи към тях и т. н.;
- как е проведено оценяването им от комисиите за проверка;

- какви характеристики са получили при анализа с методите на IRT описани във втора глава.

Направено е проучване какви са възможностите на колеги в страната, които да проведат със своите ученици пробно НВО. Каква част от учебния материал ще бъде взет до средата на месец май, за да не бъде включено неизучено учебно съдържание.

Основният етап на експеримента премина през следните стъпки:

- Създадени са 20 задачи с ИО и 3 задачи с разширен СО по новия модел на НВО за 2024, като са отчетени изводите, получени в предварителния етап. Някои от задачите с ИО са почти идентични с даваните през годините, докато в други задачи са направени промени, които отразяват откритите в предните глави на дисертационния труд проблеми. Целта е да се проследи доколко тези промени са въздействали на измерителните качества на проблемните задачи – дали са довели до подобрения или не. Някои промени са във формулировката на задачата, а други – в избора на дистракторите. Условиата на тези задачи са в приложение 7;
- Пробното НВО бе разделено на две части:
 - Първа част, в която учениците работят 75 минути, съдържа задачи само с 4 избираеми отговора, само един от които е верен;
 - Втора част, в която времето за работа е 90 минути, се състои от 3 задачи с разширен СО, на които учениците трябва да напишат пълните решения с необходимите обосновки и чертежи;
- Задачите с ИО, според модела от 2024 г., се оценяват с 2, 3 или 4 точки, като общият им сбор трябва да е 65 точки. Възможностите за разпределяне на броя на задачите според това условие са няколко, като за експеримента бе избран вариант с 3 лесни задачи (по 2 точки), 9 оптимално трудни задачи (по 3 точки) и 8 трудни задачи (по 4 точки);
- Трудността на създадените задачи с ИО бе съобразена според трудността на задачите от миналите години, които служеха за образец. В този смисъл може да се приеме, че те са апробирани;
- Задачите с разширен СО са три на брой, като едната е алгебрична, другата е с приложен характер, а третата е геометрична. Всяка от тях е с две или три подточки;

- За задачите с разширен СО са подготвени два варианта на ръководство за оценяване:
 - *традиционен вариант* – който е приет за проверка в настоящите НВО. При него с 1 точка се оценява всеки етап от решението на задачата, като е възможно за части от тези етапи да се оценяват и с дробни – 0,5 или 0,25. По този начин точките за всяка от тези задачи е 12 точки за задачи № 21 и № 23 и 11 точки за задача №22. Общият им сбор е 35 точки;
 - *експериментален вариант* – при който всяка подточка на всяка задача се разглежда като отделна задача и се оценява с цяло число точки по описания във втора глава начин, според броя на важните стъпки в нейното решаване, за да може да се приложат методите на IRT. Така 0-1 задача е № 23 А), 0-1-2 задачи са № 21 А), 21 Б), 21 В), 22 А), 22 Б) и 23 В), а 0-1-2-3 задача е № 23 Б);
- За всеки ученик е предвиден формуляр, в който да отбележи верния според него отговор на всяка задача с ИО, а проверяващият трябва да нанесе точките, получени при верен отговор на тези задачи. Задачите с разширен СО се описват на отделни листове, като получените точки за всяка подточка проверителят нанася в същия формуляр;
- Пробният изпит се проведе с ученици от различни училища в страната – основни, средни и математически гимназии в периода 27 април – 1 юни 2024 г.;

Вид училище	Град	Брой	Общо
МГ	Плевен	32	80
МГ	Монтана	25	
МГ	Разград	23	
СУ	Плевен	32	90
СУ	София	19	
СУ	Пловдив	39	
ОУ	Благоевград	38	317
ОУ	Плевен	117	
ОУ	Добрич	32	
ОУ	Монтана	67	
ОУ	Шумен	63	
Общ брой			487

Таблица 3.1. Брой на учениците, участвали в експеримента

- Броят на учениците според училището, в което учат е даден в таблица 3.1. Както се вижда, повечето ученици са от средни или основни училища, като има и от математически гимназии, където се предполага, че подготовката по математика е на

по-високо ниво от останалите училища. Участие взеха ученици от 9 административни области на България;

- Оценяването според традиционния вариант бе направено от учителите на учениците, след което работите на учениците са изпратени на автора на дисертационния труд за оценяване според експерименталния вариант. За целта тези работи са или сканирани, или изпратени по куриер.

Заклучителният етап на изследването включва:

- Анализ на задачите след провеждането и оценяването на пробното НВО;
- Сравняване по различни показатели на традиционния и експерименталния начин на точкуване;
- Предложения за начина на точкуване при бъдещи НВО.

Анализ на получените резултати след провеждане на експеримента

Точки	№	Трудност	gpbis	Тълкуване
2 т.	13	0.86	0.36	мн. лесна
3 т.	4	0.81	0.41	лесна
3 т.	10	0.81	0.26	лесна
3 т.	12	0.79	0.46	лесна
2 т.	5	0.75	0.35	лесна
3 т.	1	0.7	0.4	оптимална
2 т.	2	0.69	0.38	оптимална
3 т.	17	0.68	0.41	оптимална
4 т.	14	0.65	0.4	оптимална
3 т.	15	0.64	0.37	оптимална
3 т.	18	0.63	0.55	оптимална
3 т.	11	0.6	0.39	оптимална
4 т.	16	0.6	0.45	оптимална
4 т.	20	0.56	0.16	оптимална
3 т.	3	0.55	0.4	оптимална
4 т.	7	0.55	0.45	оптимална
4 т.	9	0.52	0.54	оптимална
4 т.	8	0.44	0.46	оптимална
4 т.	6	0.4	0.25	трудна
4 т.	19	0.39	0.42	трудна

Таблица 3.3. Наредба на задачите с ИО от 2024 г. според трудността им

№	Трудност	gpbis	Тълкуване
20	0.56	0.16	ниска
6	0.4	0.25	средна
10	0.81	0.26	средна
5	0.75	0.35	добра
13	0.86	0.36	добра
15	0.64	0.37	добра
2	0.69	0.38	добра
11	0.6	0.39	добра
1	0.7	0.4	добра
3	0.55	0.4	добра
14	0.65	0.4	добра
17	0.68	0.41	мн. добра
4	0.81	0.41	мн. добра
19	0.39	0.42	мн. добра
16	0.6	0.45	мн. добра
7	0.55	0.45	мн. добра
12	0.79	0.46	мн. добра
8	0.44	0.46	мн. добра
9	0.52	0.54	мн. добра
18	0.63	0.55	мн. добра

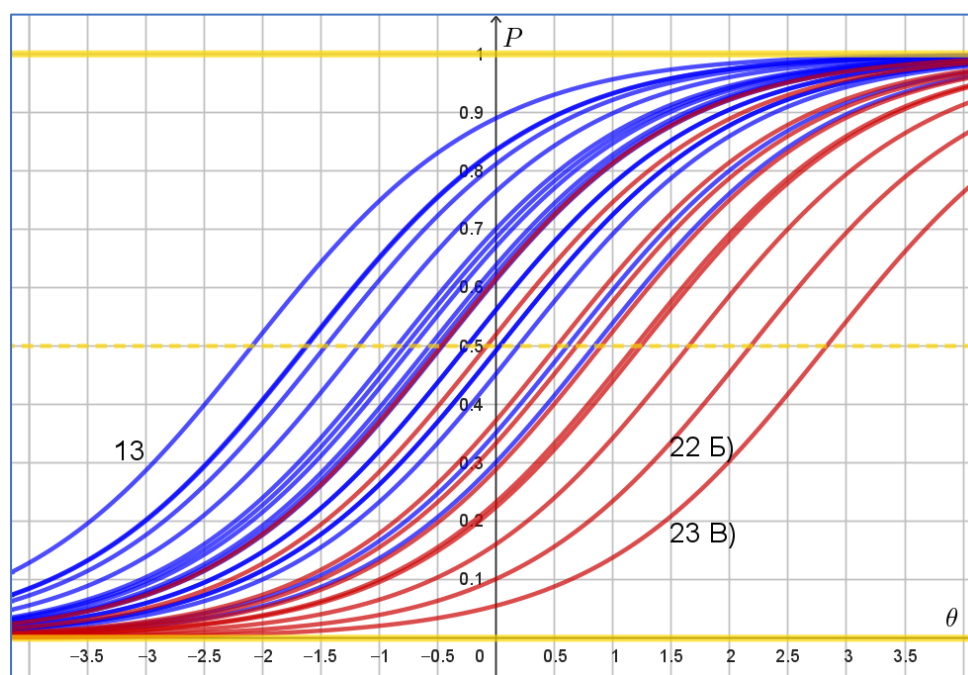
Таблица 3.4. Наредба на задачите с ИО от 2024 г. според дискриминацията им

Във втората точка е направен анализ на получените резултати. В първа подточка са анализирани задачите с ИО посредством КТТ. Установено е, че задачите от пробното

НВО съответстват по основните характеристики на тези от проведените в изследваните четири години НВО.

В таблици 3.3 и 3.4 са дадени номерата на задачите, подредени съответно според коефициентите на тяхната трудност и на дискриминативна сила ($gpbis$). Както се вижда, има разнообразие в трудността на задачите. Има една много лесна задача, 4 лесни, 13 оптимални и 2 по-трудни, което е добра характеристика ако НВО се разглежда като нормативен тест. При съставянето на теста задачите по-лесните задачи са оценени с 2 точки, трудните – с 3 точки, а най-трудните – с 4 точки.

Според таблица 3.3, има малко разминаване при по-леките задачи, но може да се каже, че то не е съществено. Дискриминацията на задачите, с изключение на №20 е на добро ниво. Тази задача е дадена, без да е апробирана. В нейното условие е заложен един експеримент – верният отговор е елементарен сбор от дадените две числа в условието и може да се уцели лесно по този начин. По-нататък ще се види, че това е доказано с методите на IRT. Задачи 6 и 10, въпреки че са апробирани, са показали по-ниска разграничителна сила.



Фиг. 3.3. Характеристични криви на задачите от 2024 г.

Във втора подточка е приложен Раш-модела върху всички задачи, като всяка подточка на задачите с разширен СО се разглежда като самостоятелна единица. Характеристичните криви (фигура 3.3) на тези задачи покриват равномерно широк интервал на трудност.

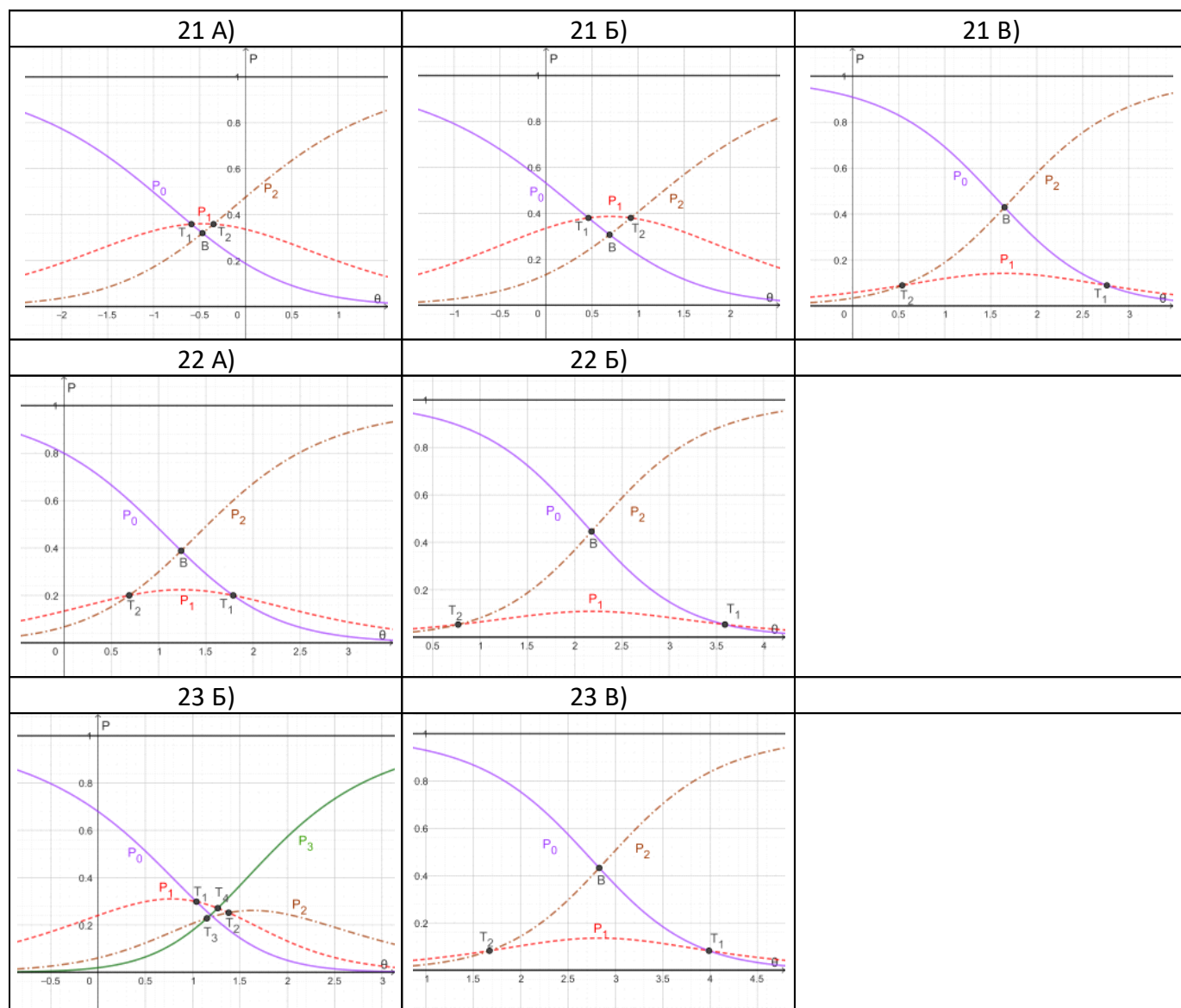
Според диаграмата на фигура 3.4 може да считаме, че тестът, съставен от представените задачи, е добре балансиран за съответните ученици, защото двете разпределения имат връх в близки стойности и покриват почти едни и същ интервал на способност/трудност. Все пак, за по-силните ученици с постижения над 2,8 липсват задачи с такава висока трудност.

Брой ученици																				
θ/b	-2.4	-2.0	-1.6	-1.2	-0.8	-0.4	0.0	0.4	0.8	1.2	1.6	2.0	2.4	2.8	3.2	3.6	4.0	4.4	4.8	
№ задача	13 4 12 5 2 11 3 23A) 6 22A) 21B) 22Б) 23B)																			
	10 1 17 16 7 8 19																			
	14 20 9 21Б) 23Б)																			
	15																			
	21A) 18																			

Фиг. 3.4. Диаграма на задачите и респондентите от 2024 г.

На фигура 3.5 може да се забележи, че само при задачи 21 А) и 21 Б) е най-вероятно учениците да получат 0, 1 или 2, както е предвидено при оценяването. При тези задачи абсцисите на точките Т₁ са по-малки от тези на Т₂. За разлика от тях, при задачи 21 В), 22 А), 22 Б) и 23 В) е по-вероятно учениците да получат 0 или 2 точки, т. е. или да решат вярно задачата, или да не я решат. При тези задачи абсцисите на точките Т₁ са по-големи от тези на Т₂. Подходящо е те или да се точкуват с 0-1 точки, или да се помисли за оценяване, при което да може да се дава междинната 1 точка при по-леки условия. Задача 23 Б) е оценявана с 0-1-2-3 точки, но както се вижда от графиката вероятността да има ученици с 2 точки е ниска. При нея е най-вероятно да има ученици с 0, 1 или 3 точки. Тя е подходящо да бъде 0-1-2 задача. Тези задачи са създадени като се има предвид точкуването на проведените НВО през изследвания четиригодишен период. Въпреки това, само при 2 (евентуално при 3) от тях има възможност да се получат междинни точки. При другите задачи учениците, ако са започнали вярно задачата, успяват да я довършат. Една от причините може да се потърси в това, че по втория модул не работят всички ученици. Повечето, които пишат решения на задачите с разширен СО са по-

подготвени и успяват да ги решат докрай. Възможно е да допуснат някоя малка грешка, но това не бива да се отразява съществено на техните точки.



Фиг. 3.5. Графики на характеристичните криви на задачите със СО от 2024 г.

Сравняване по различни показатели на традиционния и експерименталния начин на точкуване

Нека се припомним, че двата начина за оценяване, които разглеждаме тук са:

- „традиционен“ – всяка задача с ИО се оценява с 2, 3 или 4 точки, като общия сбор 65, а задачите със СО се оценяват с 11 или 12 точки – общо 35 точки. Така максималният бал от двете части на НВО е **100 точки**;
- „експериментален“ – всяка задача с ИО се оценява с 1 точка, а на всяка подточка на задачите със СО се дават максимално 1, 2 или 3 точки. Така максималният бал е **36 точки**.

За да се сравнят получените резултати по двата начина – традиционния и експерименталния, може да се подходи по различни начини. Един от тях е описан в (Банков, 2012). За целта, ще приравним точките, получени от **експерименталното (X)** оценяване (от 0 до 36 т.) към тези от **традиционното (Y)** оценяване (от 0 до 100 т.). Това се прави с формулата:

$$X_i^* = \frac{\sigma_Y}{\sigma_X} (X_i - \mu_X) + \mu_Y,$$

където

- X_i са точките на i -тия ученик според експерименталното оценяване;
- X_i^* са тези точки, приравнени към традиционното оценяване;
- μ_X и μ_Y са средните стойности на суровите балове на учениците, получени съответно от експерименталното и традиционното оценяване;
- σ_X и σ_Y са техните съответни средноквадратични отклонения.

Така се получава формулата:

$$X_i^* = 2,76(X_i - 17,63) + 49.$$

След това е получен коефициентът на корелация между **получените балове (X^*)** и тези от традиционното оценяване (Y). Стойността на този коефициент е **0,99192**, което означава, че двете оценявания са идентични. За сравнение, в таблица 3.14 са дадени точките на първите 20 ученици по трите скали. Очевидно в повечето случаи разликите са 0 или 1 точки, макар че има и по-големи разлики – от 10 точки.

Една от причините за големите разлики е когато единият и другият оценител не са оценявали еднакво напредъка на ученика в решенията на задачите със СО. Това не е необичайна ситуация при НВО и по тази причина там има арбитър, който преценява коя от оценките на двамата оценители е по-близка до реалната. Следва да се отбележи, че 1 т. по експерименталното оценяване съответства приблизително 3 т. по традиционното. Колегите, които са оценявали по традиционния начин своите ученици, понякога са склонни да дават 1-2 „бонус“ точки на ученици, които обикновено работят добре в час, но в случая са допуснали грешки. Тези грешки, обаче при експерименталното оценяване са наказвани, така както е предвидено в ръководството за оценяване.

№	X	X*	Y
1	34	94	94
2	27	75	80
3	32	89	88
4	26	72	66
5	27	75	79
6	32	89	89
7	26	72	65
8	33	92	93
9	21	58	61
10	33	92	98,5
11	35	97	94
12	34	94	95
13	24	67	77
14	28	78	82
15	30	83	85
16	36	100	94
17	36	100	100
18	21	58	64
19	34	94	96
20	9	25	24

Таблица 3.14. Точки на първите 20 ученици по три скали за оценяване

Друга ситуация, при която се получават по-големи разлики в точките между X^* и Y може да се илюстрира с ученик номер 13. Той е допуснал 1 грешка на задача 2, при което сборът на суровия му бал на задачите с ИО е $X_{13} = 19$, което дава приравнен бал $X_{13}^* = 53$, докато балът по традиционното оценяване е $Y_{13} = 63$. Така разликата между двете оценявания в общия му бал също е 10 точки.

Въпреки, че ги има, случаите на по-големи разлики са малко на брой и не влияят съществено на общата статистика, което се доказва от високия коефициент на корелация между баловете X^* и Y .

Изводи от Трета глава

В четвъртата точка са направени изводи от проведения педагогически експеримент. Резултатите от направения експеримент могат да се класифицират според трите основни посоки на изследването:

1. По отношение на задачите с избираем отговор:
 - а. Важен елемент при съставянето им е да не се допуска верният отговор да се получава чрез елементарни аритметични действия на числата в условието – задачи 12 и 20;

- б. Чертежите към геометричните задачи да не дават прекалено много информация за получаването на отговора, особено в случаи, когато е необходимо да се приложат някои свойства на обектите – задачи 17 и 19;
- в. Ако чертежът на геометричната задача е достатъчно подсказващ, да се избера отговори, които не биха могли да се налукат лесно – задача 19;
- г. Понятия, които рядко се ползват в VII клас често не водят до добра дискриминация на учениците – *масшаб* в задача 4 и *прости числа* в задача 6.

2. По отношение на задачите с разширен СО:

- а. Положителна тенденция е те да се разделят на няколко подточки, като мнението на автора е, че всяка подточка може да подпомага решенията на следващите, а не да бъде задължително независима от тях. Така по-слабо подготвените ученици имат по-голяма възможност да получат точки и във втората част на НВО – проблем, който се вижда всяка година;
- б. Всяка подточка да се оценява, като се отделят няколко важни стъпки в решението ѝ. Тези стъпки не бива да са повече от 2-3. Така има възможност да се направи изследване с помощта на вероятностното моделиране и да се сравняват по-точно резултатите от различни НВО. Да се дава възможност за сравнително лесно получаване на 1 точка, което да е сравнимо с решаване на задача с ИО.

3. По отношение на балообразуването:

- а. Оценяването на задачите с ИО с 2, 3 или 4 точки не е много оправдано. Очевидно, дори те да са добре познати и апробирани, не винаги на по-лесните задачи съответстват по-малко точки;
- б. Теорията препоръчва оценяването на задачите с ИО да е с 0-1 точки, независимо от тяхната трудност. Този коефициент показва всъщност каква част от учениците са посочили верния отговор, така че се очаква, ако един ученик е добре подготвен, той да може да посочи повече верни отговори;
- в. Оценяването на отделните подточки на задачите с разширен СО да е с цяло число точки 0-1-2 или 0-1-2-3, които да съответстват на важните стъпки в решението им;
- г. Високата стойност на коефициента на корелация между двата вида оценяване, получен в точка 3.3 на тази глава показва идентичност на двата начина на

оценяване, като експерименталният начин е податлив на анализ с методите на съвременната теория на тестовете;

- д. При нужда от унифицирана скала, може получените балове да се представят в проценти от 0 до 100.

Използването на общите задачи между пробното НВО от 2024 г. и изминалите НВО могат да послужат при бъдещи изследвания като “anchor” задачи за по-обективно сравняване на резултатите на учениците от четирите изследвани години. Този метод е приложен в точка 3 на втора глава. Чрез подобни примерни НВО, които да се проведат в специално подбрани училища в страната може да се създаде система за сравняване на тези резултати. Може да се препоръча на авторите на бъдещите НВО по математика да апробират задачите, като използват идеите на този експеримент. По този начин биха се избегнали някои неволни грешки както в условията, така и в отговорите на задачите с ИО. Също така, може да се види доколко важните стъпки в решаването на задачите с разширен СО добре разделят решението, така щото учениците да имат възможност да получават с добра вероятност междинните точки на тези задачи.

Изводите от тази глава, както и тези от втора глава потвърждават хипотезата на педагогическия експеримент. При условие, че оценяването на НВО се опрости, както е описано тук в изводи 9 и 10, по-прецизно могат да се анализират всички задачи от НВО чрез средствата на IRT. При това, тази теория е по-надеждна при сравняването на различни популации отколкото общоприетото им сравняване посредством средния успех.

Заклучение

Настоящата разработка е в резултат на дългогодишното участие на автора в комисии за съставяне и проверяване на различни изпити и оценявания на регионално и национално ниво. Фокусът в дисертационния труд са Националните външни оценявания по математика, които се провеждат с ученици, завършващи VII клас. В тази връзка са потърсени и са намерени отговори на въпроси, които от самото начало на провеждането на външните оценявания са на дневен ред – как да се тълкуват постигнатите от учениците резултати; в каква степен учениците са усвоили учебния материал; съответстват ли даваните задачи на поставените предварителни изисквания относно съдържателната и познавателната област; доколко тези задачи са „добри“ от гледна точка на теориите за тестови задачи с избираем или свободен отговор; как могат да се подобрят някои от

тях, които са показали определен „дефект“; доколко използването на НВО за вход в гимназиите е оправдано; възможно ли е оценяването на задачите да се постави на друга основа; в каква степен обучението от разстояние в електронна среда е въздействало върху постиженията на учениците?

Основна цел в дисертационния труд е да се приложат различни методи за анализ на резултати от стандартизирани оценявания на ученици по математика (НВО) и да се анализират положителните и отрицателните страни на всеки от тях. Използвайки този анализ, да се дадат препоръки за тяхното използване. Основната хипотеза е, че правилното използване на Класическата теория на тестовете, съчетана със съвременната Item responsibility theory, дава по-обективна, по-точна и надеждна картина за постиженията по математика на учениците от VII клас. Тази хипотеза е проверена в дисертационния труд.

В подкрепа на тази теза авторът е проучил и анализирал публикации, свързани с прилагането на двете теории в национални и международни изследвания на ученическите постижения. Изучил е различни софтуерни продукти за обработка на резултатите от проведените тестове.

От проведените наблюдения и анализи може да се направят следните **изводи**:

От първа глава:

- Застъпени са всички теми от съдържателната рамка на тестовата спецификация;
- Задачите с ИО основно са на познавателно равнище Знание и Разбиране
- Трудността и дискриминативната сила са с добри показатели, с няколко малки изключения;
- Задачите със СО не дават добавена стойност при голяма част от учениците в малките и средни по големина области на страната. Това, обаче за София-град и отчасти за Пловдив не е така;
- COVID епидемията се отрази на резултатите на учениците, особено на тези, които се обучаваха в ОРЕС и завършиха през 2022 година VII клас;

От втора глава:

- Изследването с различните модели на IRT показва, че двупараметричният модел се съгласува най-добре с емпиричните данни. Резултатите от този модел потвърждават изводите, получени с КТТ;

- Прилагането на трипараметричния модел дава възможност да се откриват задачи, които имат висок коефициент на налучкване. За тези задачи е добре да се потърси по-добър вариант или да не се включват в следващи оценявания;
- Трудността на задачите с ИО се съгласува добре със способностите на учениците, като липсват задачи, които са за много силните и много слабите ученици;
- Сравняването на постиженията на учениците с помощта на anchor-задачи показва, че те се понижават през изследвания период от четири години, като това понижение се оказва статистически значимо;
- Точкуването на задачите по общоприетия начин не винаги съответства на трудността им. Има не малко примери, в които трудна задача се оценява с малко точки, а лесна – с повече точки;
- Показана е идея за оценяване, базирана на резултатите на учениците, която може да бъде внедрена в бъдещи НВО;

От трета глава:

- Установени са няколко причини, които могат да влошат качествата на задачите с ИО, като са предложени начини откриване на такива задачи, както и варианти за подобряването на тези качества;
- Оценяването на задачите с ИО е добре да е с 1 точка при верен отговор и 0 точки в останалите случаи;
- Оценяването на задачите със СО не е нужно да бъде раздробено, както се прави в настоящите НВО. Достатъчно е техните подточки да бъдат оценявани с 1, 2 до 3 точки, които съответстват на важните стъпки в решаването им;
- При нужда от унифицирана скала от 0 до 100, суровите балове на учениците, получени от горните две оценявания, могат да се представят в процентна скала.

Приноси на дисертационния труд

Научноприложни приноси

1. Съпоставени са различни модели на IRT (с един, два и три параметъра) с получените емпирични данни върху задачите с ИО от проведените НВО в периода 2020-2023 г., като са изследвани резултатите на всички ученици от четири административни области в България. Обосновано е, че двупараметричния модел съответства най-добре на тези данни.

2. Направена е съпоставка на различни методи за оценяване на задачите със СО. Създадена е опростена скала за оценяване на задачите със СО, така че да може да се приложат методите на IRT за изследване на техните показатели. Тази скала е обоснована на базата на получените резултати от проведените НВО;
3. Сравнени са постиженията на учениците по математика в изследваните четири години, като е приложен метод, базиран на двупараметричния модел на IRT, за сравняване на различни популации чрез така наречените anchor-задачи. Установено намаляване на тези постижения с всяка следваща година;
4. Предложени са варианти за трансформация на точките от задачите с ИО в широко използваната у нас шестобална система;

Приложни приноси

5. Направен е изводът, че приложените методи на КТТ върху задачите с ИО от НВО след VII клас в изследвания четиригодишен период са добре балансирани по отношение на коефициентите на трудност и разграничителна сила.
6. Показано е, че задачите с ИО от НВО в този период покриват изискванията по отношение на съдържателната и познавателната област от рамката за оценяване;
7. Установено е, че задачите със СО от НВО в изследвания период не допринасят съществено към балообразуването с цел класиране на учениците;
8. Проведен е педагогически експеримент, като са съставени 23 авторски задачи, които са съобразени с тези на проведените НВО в изследвания четиригодишен период;
9. Установено е, че има разминаване в наредбата по трудност на задачите с ИО според общоприетия начин (с 0, 2, 3 или 4 точки) и според условието за вярна-невярна задача (с 0 или 1 точка);
10. Показан е метод за оценяване на задачите с разширен СО според броя на „важните“ стъпки в решаването им;
11. Установено е, че има отлично съответствие на получените балове на учениците по двата начина, наречени традиционен и експериментален.

Насоки за бъдещо развитие

Някои идеи за бъдещи разработки, които се появиха в процеса на разработката на този труд:

- ✓ Използването на общите задачи между проведеното пробното НВО от 2024 г. и миналите НВО в периода 2020-2023 година могат да послужат като anchor-задачи за по-обективно сравняване на резултатите на учениците от четирите изследвани години. Този метод е приложен в точка 3 на втора глава;
- ✓ Чрез подобни примерни външни оценявания, които да се проведат в специално подбрани училища в страната, може да се създаде система за сравняване на резултати от проведени такива оценявания по математика. Същото може да се експериментира и в други учебни дисциплини;
- ✓ Да се проучи въпросът с обективността на оценяването и доколко прилагането на методите за изследване на тестове с IRT биха подобрили тази обективност.

Литература

- Алашка, Р. (2016). Двухфакторни вероятностни модели в образованието. *Механика, транспорт, комуникации*, 14(3/2), VII-7 - VII-12.
- Алашка, Р. (2017). *Приложение на вероятностни модели за анализ на резултати от изпити и тестове, Дисертационен труд за присъждане на научната и образователна степен „доктор“*. София: Софийски университет "Св. Кл. Охридски".
- Банков К., Витанов Т. (2010). Външното оценяване по математика, състояние и перспективи. *Математика и математическо образование*. Албена: СМБ.
- Банков, К. (2002). Вероятностно моделиране за измерване на ученическите постижения. *Математика и информатика*(4), 41-49.
- Банков, К. (2012). *Увод в тестологията*. София: Изкуства.
- Банков, К. (2024). *Широко мащабни оценъчно-диагностични педагогически изследвания*. София: Университетско издателство "Св. Климент Охридски".
- Банков, К. (2023). Точкуване на задачи и използване на скали при стандартизирани оценявания. *Математическото образование 75-годишна мисия и история*, 73-86.
- Бижков Г., Краевски В. (2002). *Методология и методи на педагогическите изследвания*. София: Университетско издателство "Св. Кл. Охридски".
- Бижков, Г. (1996). *Теория и методика на дидактическите тестове*. София: Просвета.

- Данчев, П., К. Банков, В. Стоименова и др. (2015). *Апробация на статистически модели за оценка на добавената стойност на училищата, използващи резултати от национални оценявания. Основни резултати и изводи*. Washington, D.C.: World Bank Group.
- Джалев, Л. (2013). *Сравнителен анализ на приложимостта на две психометрични теории на тестовете (върху данни от теста за общообразователна подготовка). Дисертационен труд за присъждане на научната и образователна степен „доктор“*. София: Нов български университет.
- Джалев, Л. (2014). Приложимост на Класическата тестова теория и Теорията за отговор на тестов въпрос: преглед на литературата по въпроса. *Българско списание по психология*(1-3), 81-102.
- Иванов, И. (2006). *Педагогическа диагностика*. Шумен: Университетско издателство „Епископ Константин Преславски“.
- Ким, В. С. (2007). *Тестирование учебных достижений. Монография*. Уссурийск: УГПИ.
- МОН-2020. (16.09.2024 г.). НВО за учениците в VII клас през учебната 2019-2020 година. Извлечено от МОН: https://www.mon.bg/nfs/2020/06/nvo-viikl-math_17062020.pdf
- МОН-2021. (16.09.2024 г.). НВО за учениците в VII клас през учебната 2020-2021 година. Извлечено от https://www.mon.bg/nfs/2021/06/nvo-math_7kl_18062021.pdf
- МОН-2022. (16.09.2024 г.). НВО за учениците в VII клас през учебната 2021-2022 година. Извлечено от https://www.mon.bg/nfs/2022/06/7kl_nvo_math_16062022.pdf
- МОН-2023. (16.09.2024 г.). НВО за учениците в VII клас през учебната 2022-2023 година. Извлечено от https://www.mon.bg/nfs/2023/06/nvo_math_7klas_variant2_16.06.2023.pdf
- Портал за отворени данни на Република България. (16.09.2024 г.). Извлечено от <https://data.egov.bg/>
- Стоименова, Е. (2000). *Измерителни качества на тестове: В помощ на авторите на тестове*. София: БАН. Институт по математика и информатика.
- Стоянова, Ф. (1996). *Тестология за учители*. София: Атика.
- Цонев, П. (2022). Някои изводи върху резултатите от националното външно оценяване по математика за VII клас. *Математика и информатика*, 65(6), 587-601.
- Цонев, П. (2023-а). Избор на подходящ модел за изследване на задачите от НВО по математика след VII клас според методите на IRT. *Математика и информатика*, 66(4), 491-505.
- Цонев, П. (2023-б). Сравнение на резултатите от задачите с избираем отговор от НВО по математика след VII клас, проведени през 2020, 2021 и 2022 години. *Годишна международна научна конференция на ВВВУ „Георги Бенковски“*, (стр. 353-363). Долна Митрополия.

- Цонев, П. (2024-а). Примерно точкуване на задачите със свободен отговор от Националното външно оценяване след VII клас през 2023 година, базирано на вероятностно моделиране. *Математика и математическо образование*, 53, 161-168.
- Цонев, П. (2024-б). Анализ на някои задачи с избираем отговор от пробно национално външно оценяване, проведено през 2024 година. *Годишна международна научна конференция на ВВВУ „Георги Бенковски“*, (стр. 380-389). Долна Митрополия.
- ЦОПУО. (16.09.2024 г.). Извлечено от Център за оценяване в предучилищното и училищното образование: <https://www.copuo.bg>
- Чехларова, Т. (2022). Национално външно оценяване по математика в IV клас през 2022 година. *Математика и информатика*, 65(6), 344 – 357.
- Янкова, Т. (16 09 2024 г.). 21. ПРОВЕРКА ЗА НОРМАЛНОСТ НА ГС. Извлечено от Упражнения по Вероятности и Статистика – Таня Янкова: https://www.fmi-plovdiv.org/evlm/DBbg/database/probstat/VerStatXPtema21_Testing%20normality%20_BG.pdf
- Birnbaum, A. (1968). *Some latent trait models and their use in inferring an examinee's ability* (Lord, F.M. and Novick, M.R., Eds., *Statistical Theories of Mental Test Scores* изд.). Addison-Wesley: Reading, 397-479.
- GeoGebra. (16.09.2024 г.). Извлечено от <https://www.geogebra.org>
- Hambelton, R., Swaminathan, H., Rogers, H. (1991). *Fundamentals of item response theory*. SAGE Publications.
- Lord, F. M., M.R. Novick (ed.). (1968). *Statistical theories of mental test scores*. Reading/Mass, Addison-Wesley.
- Meyer, J. P. (16 09 2024 г.). Извлечено от Psychomeasurement Systems: <https://itemanalysis.com>
- Rasch, G. (1980). *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.
- Tsonev, P. (2024). Challenging Mathematic Task-Types for Students Finishing 7th Grade in Bulgarian Schools. *KNOWLEDGE-BASED ORGANIZATION*, 30 (Issue 2), стр. 1-4. Sibiu.

Забележка:

В литературата са включени всички заглавия, които са цитирани в дисертацията. Тези от тях, които не са цитирани в автореферата са с наклонен шрифт. Броят на цитираните заглавия в автореферата е 27.

Публикации на автора, свързани с темата на дисертацията

- Цонев, П. (2023). Сравнение на резултатите от задачите с избираем отговор от НВО по математика след VII клас, проведени през 2020, 2021 и 2022 години. *Годишна международна научна конференция на ВВВУ „Георги Бенковски“*, (стр. 353-363). Долна Митрополия.
- Цонев, П. (2024-а). Примерно точкуване на задачите със свободен отговор от Националното външно оценяване след VII клас през 2023 година, базирано на вероятностно моделиран. *Математика и математическо образование*, 53, 161–168.
- Цонев, П. (2024-б). Анализ на някои задачи с избираем отговор от пробно национално външно оценяване, проведено през 2024 година. *Годишна международна научна конференция на ВВВУ „Георги Бенковски“*, (стр. 380-389). Долна Митрополия.
- Tsonev, P. (2024). Challenging Mathematic Task-Types for Students Finishing 7th Grade in Bulgarian Schools. *International conference KNOWLEDGE-BASED ORGANIZATION*, 30(2) 1-4.

Декларация за оригиналност

от Павлин Иванов Цонев

докторант в задочна форма на обучение

към докторска програма „Методика на обучението по математика и информатика“ -

Методика на обучението по математика

при ФМИ на СУ „Св. Климент Охридски“

Във връзка с провеждането на процедура за придобиване на образователната и научна степен „доктор“ във ФМИ на СУ „Св. Климент Охридски“ и защита на представения от мен дисертационен труд, декларирам:

Резултатите и приносите на проведеното дисертационно изследване, представени в дисертационния ми труд на тема „Методи за анализ на резултати от стандартизирани оценявания на ученици по математика“ са оригинални и не са заимствани от изследвания и публикации, в които нямам участия.

София, 2025 г.

Декларатор:

Павлин Цонев

Благодарности

Искрено благодаря на:

- научния ми ръководител проф. Банков за неговия професионализъм и търпение към потока от моите непрекъсващи въпроси. За куража, който ми даде при разработката на този труд и публикациите, които го съпътстваха. За неговите ясни и конкретни напътствия, които идваха в точния момент;
- колегите от ЦОПУО и конкретно неговия директор г-жа Неда Кристанова, които ми съдействаха да получа данните от проведените НВО по математика за VII клас, без които това изследване би било невъзможно;
- учителите, с които си сътрудничих, за да проведе педагогическия експеримент:
 - от Благоевград – Медиха Топалова;
 - от Добрич – Милена Аврамова;
 - от Монтана – Йорданка Еленкова и Стефчо Наков;
 - от Плевен – Ани Цонева, Валери Наумов, Георги Мелниклийски, Диана Данова, Долорес Ненова, Елка Линкова, Емилия Мелниклийска, Маруся Кехайова, Миланка Дечева, Петранка Кубратова, Пламен Митев, Пламен Петранов, Светла Петрова, Силвия Спиридонова, Теодора Тодорова;
 - от Пловдив – Елена Тодорова;
 - от Разград – Емилия Станкова;
 - от София – Димитър Димитров;
 - от Шумен – Антоанета Ковачева;
- колегата Ваня Кацарска от ВВВУ „Г. Бенковски“, която ми оказва неоценимо съдействие в английския вариант на реферата;
- семейството и колегите ми от ВВВУ „Г. Бенковски“ за проявеното разбиране и морална подкрепа!