

ТЕСТОВИТЕ ЗАДАЧИ ОТ ДЪРЖАВНИЯ ЗРЕЛОСТЕН ИЗПИТ ЗА ПРОФИЛИРАЩ УЧЕБЕН ПРЕДМЕТ „ИНФОРМАТИКА“ ПРЕЗ УЧЕБНАТА 2021/2022 ГОДИНА

Доц. д-р Димитър Атанасов¹⁾, д-р Красимир Манев²⁾,
доц. д-р Весела Стоименова³⁾, държавен експерт Ралица Войнова⁴⁾

¹⁾ Департамент „Информатика“ – Нов български университет

²⁾ Сп. „Математика и информатика“

³⁾ Факултет по математика и информатика – Софийски университет

⁴⁾ Министерство на образованието и науката

Резюме. През учебната 2021/2022 година завърши първият випуск зрелостници от българските средни училища, изучавали профилиращ предмет „Информатика“ и се проведеха първите държавни зрелостни изпити за завършилите профила. В статията са направени някои количествени и качествени оценки на резултатите от първата част на изпита, в която учениците отговаряха на въпроси с избираем и свободен отговор. Получените резултати биха били полезни за подобряване на тестовата част на изпита при подготовката на темите за следващи изпити.

Ключови думи: профилирано обучение; информатика; държавен зрелостен изпит; статистическа оценка; въпроси с избираем отговор; въпроси с отворен отговор

1. Въведение

Няма съмнение, че българската софтуерна индустрия е измежду най-успешните отрасли на българската икономика. От ключово значение, за да бъде този отрасъл все така успешен, е подготовката на кадри за неговите нужди. При съвременните изисквания към един професионалист в коя да е област на знанието не е възможно той да бъде подготвен на необходимото за практиката ниво само в рамките на университетското образование. Тук е мястото на средното училище, което трябва да даде на учениците сериозна базова подготовка в съответната дисциплина, за да могат те да преминат успешно обучението в университетското ниво и да се подготвят добре за упражняване на бъдещата си професия.

На фона на достатъчно обемистите учебни програми в основните учебни дисциплини – български език, математика, физика, химия, биология, история, гео-

графия и т.н. – учебната програма за задължителна подготовка по информатика в средното ни училище е недостатъчна дори като се вземат предвид паралелните и пряко свързани с информатиката дисциплини „Информационни технологии“ и „Компютърно моделиране“. Затова обучението по профилиращ предмет "Информатика" е призвано да компенсира до голяма степен тази недостатъчност.

Обучението по профилиращ предмет "Информатика" започна през 2020 година. Програмата е предвидена за изучаване в две учебни години и съдържа четири модула: „Обектно ориентирано проектиране и програмиране“ (72 учебни часа), „Структури от данни и алгоритми“ (72 учебни часа), „Релационен модел на бази от данни“ (72 учебни часа) и „Програмиране на информационни системи“ (54 учебни часа). Първите два модула се изучават в XI клас, а вторите два – в XII клас.

Държавният зрелостен изпит (ДЗИ) на учениците, изучавали профилиращия предмет, се състои от теоретична част, в която учениците трябва да отговорят на 24 въпроса – 16 с избираем отговор (1 от 4) и 8 със свободен отговор. Правилният отговор на всеки от въпросите с избираем отговор се оценява с една точка, а всеки въпрос с отворен отговор се оценява с до 3 точки в зависимост от неговата вярност и пълнота. Оценяването на въпросите със свободен отговор се извършва от двама оценители, като крайната оценка е средното аритметично, при разлика от 1 точка, или след арбитраж при по-голяма разлика. Времето за решаване на теста е 90 минути.

Практическата част на изпита трае 150 минути и се състои от 4 задачи, за всяка от които изпитваните трябва да напишат програмен код, който я решава. Една от задачите е с алгоритмичен характер, а други две изискват обектно ориентирано проектиране на реални обекти и програмна реализация на проекта. Тези три задачи трябва да бъдат решени с програми, написани на един от двата разрешени езика за програмиране – C# или Java. Четвъртата задача е за създаване на неголяма база от данни и трябва да бъде решена с програма, написана на SQL.

Няма да изписваме тук въпросите от коментираната тема, за да не увеличаваме ненужно обема на статията, освен когато съдържанието на въпроса е съществено за коментирането му. Изпитните теми за двете сесии могат да бъдат разгледани от посочените в БЕЛЕЖКИ връзки към сайта на МОН за ДЗИ, проведен през май¹, и ДЗИ, проведен през август².

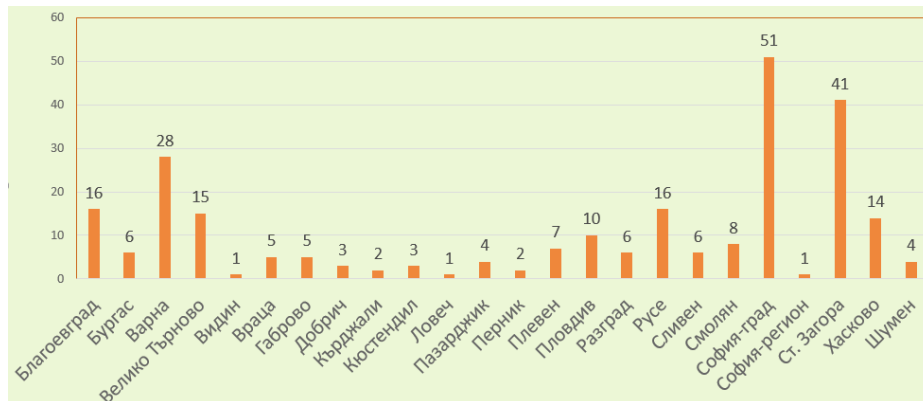
В тази статия сме направили опит да оценим както количествено, така и качествено изпитните въпроси от първата част на ДЗИ за профила „Информатика“, като се надяваме с това да помогнем за подобряване съдържанието на изпитните материали за следващите сесии на този изпит.

2. Държавните зрелостни изпити (ДЗИ) през 2022 година

През 2022 година се проведеха две сесии на ДЗИ за профилиращия предмет „Информатика“ – една през май-юни и една през септември-октомври. На се-

сията през май-юни се явиха 254 ученици, а на сесията през август-септември – 14. Тъй като броят на явилите се на втората изпитна сесия е твърде малък, за да се направят значими от статистическа гледна точка изводи, в изложението по-долу са показани и коментирани резултатите само от първата сесия.

Явилите се на първата сесия ученици са от 24 области на България. Броят на учениците за всяка от областите е показан на фиг. 1.



Фигура 1. Брой явили се зрелостници по области

Това, което прави впечатление на фиг. 1, е отсъствието на зрелостници от 4 от 28-те области, както и сравнително ниският брой явили се на ДЗИ в някои от големите градове на България.

Средният успех на участниците в сесията през май-юни е Добър 4.16, което може да се приеме като добра най-обща оценка за качеството на изпитната тема. Средният успех на участниците по области е показан на фиг. 2.



Фигура 2. Среден успех по области

Като качество, оценките се разпределят така: Отличен 6 – 20 оценки, Отличен – 40 оценки (или общо около 23.62%), Много добър – 68 оценки (26.77%), Добър – 60 оценки (23.62%), Среден – 20 оценки (7.87%), и Слаб – 46 оценки (18.11%). Общо положителните оценки са 81.89%. Процентът на слабите оценки не е малък, но обсъждането на този факт не е предмет на настоящата статия.

3. Статистическа оценка на тестовите въпроси

Използването на тестове като инструмент за оценяване способностите и знанията на индивидите е свързано с редица трудности и специфики. Основният проблем възниква при определяне на *психометричните характеристики* на теста. Те са от особена важност за определяне измерителните качества на теста и използването му за оценка на способности и знания. Изследването на този аспект на тестовете позволява коригиране и оптимизиране на тестовите въпроси с оглед да се постигнат определени предварително набелязани характеристики на тестовете. В по-късен етап възможността да се разполага с тестови задачи с предварително известни характеристики, позволява да се създават тестове, отговарящи на **предварително зададени изисквания**, както и **приравняването на резултатите**, получени при различни тестове.

Най-общо психометричните характеристики на тестовете и тестовите въпроси могат да се оценяват с два основни класа статистически модели – *класически* и *психометрични*. Класическите тестови модели са базирани основно на връзката между истинския и наблюдавания тестов бал (т.нар. *надеждност* на теста). Описание на тези модели може да бъде намерено в (Crocker & Algina 1986). Силната зависимост на този тип модели от характеристиките на изследваната популация и спецификата на теста прави използването им неособено практично от приложна гледна точка.

За преодоляването на тези недостатъци на класическите модели са въведени (и продължават да се развиват) редица психометрични модели, които представят резултата от теста като взаимодействие между индивида и тестовите задачи. Исторически, един от първите такива модели е т.нар. Rasch модел (Crocker & Algina 1986), при който вероятността $P(X=1)$ за верен отговор на дихотомен тестов въпрос трябва да удовлетворява линейната зависимост:

$\ln \frac{P(X=1)}{P(X=0)} = \theta - b$, където $b \in R$ е характеристика на въпроса, изразяваща неговата трудност, а $\theta \in R$ е характеристика, изразяваща подготвеността (знанията) на индивида. По-нататъшно развитие на този модел може да се проследи в Теорията за отговор на тестов въпрос (IRT). При него вероятността за верен отговор на дихотомен въпрос се задава чрез така наречения 3PL модел:

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

където $b \in R, a > 0$ и $c \in (0,1)$ задават съответно *трудност*, *дискриминативна сила* и *вероятност* за отгатване на правилния отговор на тестовия въпрос (Crocker & Algina 1986).

Характерна особеност на тези модели е необходимостта от относително сложни изчислителни процедури и съответен софтуер (виж например Robitzsch 2022) за определяне на неизвестните параметри. Това се обуславя най-вече от факта, че е невъзможно едновременно да се определят параметрите на тестовите задачи и на индивидите. Това налага използването на итеративни алгоритми от типа на ЕМ алгоритъм (Baker & Kim 2023).

Съществен недостатък на тези модели в тяхното практическо прилагане е, че съгласно модела подготвеността на индивидите (представени чрез θ) приема стойности върху цялата реална права. Това е нереалистично, тъй като подготвеността, която може да бъдат измерена с даден тест, е ограничена. Така възникват редица проблеми, най-вече по отношение на приравняването на резултатите от различни тестове, за решаването на които се налагат редица ограничения върху IRT модела.

Алтернативен подход е да се дефинира психометричен модел, при който оценяваната подготвеност на индивида приема стойности в ограничен интервал. Например в (Ramsay & Wiberg 2017) е представен модел, при който $\ln \frac{P(X=1)}{P(X=0)}$ се интерполира със сплайн функции в интервала от стойности на тестовия бал.

Подходът, наречен *D-scoring* (Dimitrov 2020), който сме приложили тук, има някои съществени преимущества по отношение на интерпретацията на стойностите на оценените параметри. Вероятността за верен отговор ($X = 1$) на дихотомен въпрос се задава чрез:

$$P(X_{ij} = 1 | d_i, b_j, s_j) = P(d_i, b_j, s_j) = \frac{1}{1 + \left(\frac{b_j(1 - d_i)}{d_i(1 - b_j)} \right)^{s_j}} \quad (1)$$

където $d_i \in [0, 1]$ представя подготвеността на индивид i ; $i = 1, 2, \dots, N$, като пропорция от знанията, изисквани от теста, които той покрива, $b_j \in [0, 1]$ задава параметър на локация на въпроса j ; $j = 1, 2, \dots, J$, която се интерпретира като трудност, а $s_j > 0$ задава фóрмата на кривата на тестовия въпрос. Тук с $N > 0$ е означен броят на индивидите, участващи в теста, с $J > 0$ – броят на тестовите задачи, а $X_{ij} \in \{0, 1\}$ е индикатор за коректния отговор на тестов въпрос j от индивид i . Оценяването на параметрите може да се реализира посредством стандартни библиотеки, реализирани в средите R (Atanasov 2022) и MATLAB (Atanasov 2020). Съгласно този модел:

$$\ln \frac{P(X_{ij} = 1)}{P(X_{ij} = 0)} = -s_j \left(\frac{b_j(1 - d_i)}{d_i(1 - b_j)} \right). \quad (2)$$

Този модел позволява два типа оценяване на неизвестните параметри. При т.нар. класическо оценяване трудността b_j на въпроса може да се представи като класическата трудност δ_j , свързана с пропорцията на верните отговори π_j , а именно $\delta_j = 1 - \pi_j$, $s_j = 1$. Тогава подготвеността на индивида i се представя като претеглената сума:

$$d_i = \frac{1}{\sum_{j=1}^J \delta_j} \sum_{j=1}^J X_{ij} \delta_j. \quad (3)$$

Алтернативно, параметрите на модела могат да се оценят посредством метода на *максималното правдоподобие*. По този начин се получава така нареченият *латентен модел* (Dimitrov & Atanasov 2021). Интерпретацията на тези параметри е същата, като се отчита, че получените стойности са реализации на максимално правдоподобни оценки на ненаблюдаеми характеристики.

При тези условия тестовата информационна функция $T(d)$ за дадено ниво на способностите d може да се представи като (Dimitrov & Atanasov 2021):

$$T(d) = \sum_{j=1}^J \frac{s_j P(d, b_j, s_j) (1 - P(d, b_j, s_j))}{d^2 (1 - d)^2}. \quad (4)$$

По аналогия с популярния Graded Response Model (GRM) (Samejima 1997) този модел може да се използва и за представяне на тестови задачи с повече от една степен на оценката. Нека оценката на тестовия въпрос j приема стойност от скалата $0, 1, \dots, k$, $k > 0$. Тогава за отношенията на вероятностите за преминаване от оценка $g - 1$ към оценка $g = 1, 2, \dots, k$, аналогично на (2), имаме:

$$\ln \frac{P(X_{ij} = g)}{P(X_{ij} = g - 1)} = -s_j^{(g)} \left(\frac{b_j^{(g)}(1 - d_i)}{d_i(1 - b_j^{(g)})} \right), \quad (5)$$

където $b_j^{(g)}$ и $s_j^{(g)}$ са параметрите на кривата, задаваща вероятността за получаване на оценка g , при условие че ученикът с оценявани способности d_i покрива изискванията за оценка $g - 1$. Така един тестов въпрос, оценяван с нарастваща скала, може да се разглежда като фамилия от криви на тестови въпроси, съответстващи на вероятността ученикът да бъде оценен с дадена оценка g .

В рамките на нашето изследване приложението на тази методология върху данните от теста на ДЗИ е реализирано в средата MATLAB (MATLAB 2020), използвайки пакета deltaScoring (Atanasov 2020). При въпросите с избираем отговор е използвана дихотомна скала $\{0, 1\}$, указваща дали отговорът, даден от индивида, е коректен съгласно ключа с верни отговори на теста. При въпросите

с със свободен отговор е използвана скала за оценяване $\{0, 1, 2, 3\}$. За анализа на резултатите е използвана модификацията на GRM, приложена върху оценките на първия проверител. Това не е съществено ограничение, тъй като случаите, в които оценките на двамата проверители се различават, са малък брой.

4. Коментар на статистическата оценка на тестовете

Ще разгледаме отделно групата въпроси с избираем отговор и групата въпроси с отворен отговор, тъй като характерът на статистическите оценки в двата случая е различен.

4.1. За въпросите с избираем отговор

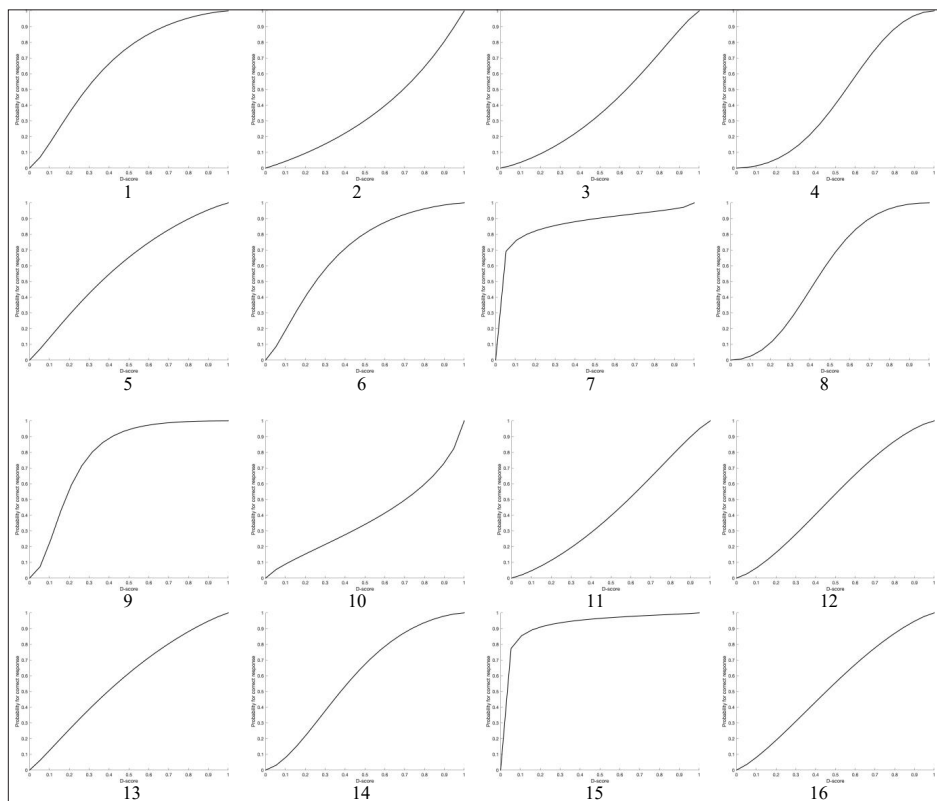
Резултатите от приложението на модела D-scoring върху данните от въпросите с избираем отговор са представени в табл. 1, където колоната δ_j съдържа класическата тестова трудност на въпроса, b_j и s_j са максимално правдоподобните оценки на латентните параметри на тестовите задачи, а $SE\ b_j$ и $SE\ s_j$ са техните стандартни грешки на оценката.

Таблица 1. Параметри на въпросите с избираем отговор

Въпрос	δ_j	b_j	s_j	$SE\ b_j$	$SE\ s_j$
1	0.1447	0.2475	1.1061	0.0596	0.2274
2	0.4736	0.6824	1.0939	0.0290	0.1820
3	0.4078	0.6256	1.1310	0.0312	0.1879
4	0.3289	0.5558	1.7524	0.0253	0.2615
5	0.2105	0.3801	1.3310	0.0449	0.2341
6	0.1315	0.1927	0.9364	0.0668	0.2129
7	0.0789	0.0201	0.5638	0.0333	0.2209
8	0.1842	0.4319	1.8821	0.0317	0.3002
9	0.0526	0.0676	0.9985	0.0503	0.2927
10	0.4736	0.6934	0.8171	0.0367	0.1512
11	0.3684	0.5941	1.1527	0.0327	0.1917
12	0.2631	0.4524	1.1639	0.0437	0.2040
13	0.2500	0.3958	1.0616	0.0520	0.1978
14	0.1842	0.3644	1.3427	0.0458	0.2383
15	0.0131	0.0381	1.1313	0.0392	0.3990
16	0.2763	0.4266	0.8475	0.0590	0.1689

На фиг. 3 са представени *характеристичните криви* на отделните тестови въпроси с избираем отговор (номерацията на отделните графики съответства на номера на теста). По хоризонталната ос на графиката със стойности от 0 до 1 е скалата, задаваща подготвеността на ученика, а по вертикалата, също със стойности от 0 до 1 – вероятността за посочване на правилен отговор при тази подготвеност.

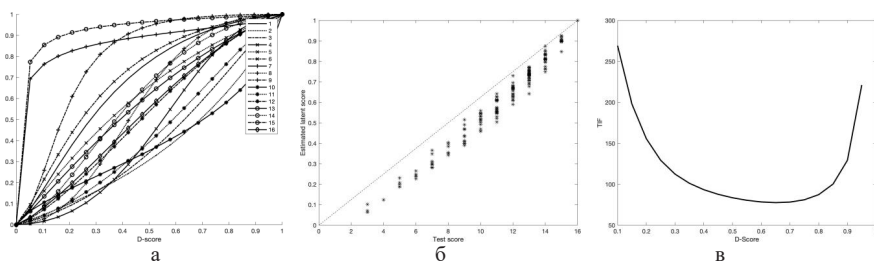
Когато кривата е разположена в горната лява половина на графиката, това показва, че при сравнително ниско ниво на подготвеност на ученика вероятността за правилен отговор е висока, съответно въпросът може да се определи като лесен. Обратно, когато кривата се намира под диагонала, в долната дясна част на графиката, това означава, че вероятността за правилен отговор при ниска подготвеност на ученика е малка, и съответно въпросът се определя като труден. Въпросите, чиито оценяващи криви са близки до диагонала на графиката, са със средна сложност.



Фигура 3. Характеристични криви на въпросите с избираем отговор

При това оценяване на въпросите, за да приемем, че една тема от избрания брой въпроси е добра, би трябвало кривите на въпросите от темата, поставени в чертожното поле, да го „запълват“ равномерно (виж фиг. 4а). Тогава тестът е добре балансиран и в него има въпроси с различна трудност.

Това, което се наблюдава, е, че преобладаващата част от задачите от тестовата част на проведения ДЗИ са със средна трудност. Две задачи са изключително лесни и на практика не допринасят за оценяване знанията на изпитваните. Третият по трудност въпрос се различава значително от първите два, а четвъртият и петият са еднакво трудни и сравнително отдалечени от третия. Останалите 11 въпроса може да бъдат оценени като средно трудни. Липсват характеристични криви в долния десен ъгъл графиката, което означава, че отсъстват относително трудни задачи. Това не позволява да се разграничат учениците с различна висока степен на знания, т.е. оценките Много добър, Отличен и Отличен б биха могли да са поставени, без да има значими разлики в подготовката на учениците, които са ги получили. Това, разбира се, не е от съществено значение при конкретната реализация на теста, но при следващи изпитвания би могло да се потърси и по-добро балансиране на темата по трудност.



Фигура 4. Обща оценка на въпросите с избираем отговор

Съответствието между наблюдавания тестов бал и оценените способности d_i, d_i на учениците е представено на фигура 4б. Наблюдава се отместване на тестовия бал от оценените способности с около 2 точки, т.е. тестовият бал надценява подготовеността на учениците. Вероятно това се дължи на наличието на две тестови задачи, които са изключително лесни и не участват в определянето на подготовеността на учениците. Добре е, че това отклонение е относително постоянно, за всички нива на оценяваната подготовеност, което означава, че няма ощетяване на едни ученици спрямо други.

На фигура 4в е представена *тестовата информационна функция*, получена чрез (4). Тя показва каква информация ни носи тестът за ученици със съответните знания. По оста x със стойности от 0 до 1 е скалата на оценените знания, а по вертикалната ос е способността на теста да оцени коректно ученик със съответните способности. Идеалната форма на графиката би била хоризонтална линия $y = c$, въпреки това, от практическа гледна точка, е удобно да се използват и криви с различна форма.

Съгласно тестовата информационна функция, тестовата част на ДЗИ е информативна за зрелостниците с най-ниски и най-високи знания. Ако ученикът е получил висок резултат, то той наистина има отлични знания. Ако пък

има нисък резултат, знанията му наистина са недостатъчни. За учениците със средни стойности на бала получената оценка не дава възможност да се направи прецизен извод за техните знания. Това вероятно се дължи на факта, че по-голямата част от задачите са средно трудни.

Нека покажем кои от тестовите с избираем отговор в темата са се оказали най-лесни и кои – най-трудни за учениците. Въпросът, който се оказва **най-лесен**, е с номер 7 ($b_j = 0.0201$). В него се изисква да се определи правилният ред на действията при записване на данни в текстов файл. Елементарността на този въпрос се определя от факта, че редът на тези действия е подчинен на елементарна логика. По-сериозен проблем с този въпрос е, че в езика C# (а вероятно така е и в Java) създаването на инстанция на класа за четене StreamReader (а не на „променлива от тип текстов файл“, както е написано във въпроса) и отварянето на файл за запис не са две отделни действия и въпросът е некоректно зададен.

Вторият **много лесен** въпрос е с номер 15 ($b_j = 0.0381$). В него е зададена структурата на таблица от релационна база от данни и се пита коя е заявката, с която да се извлекат „всички записи от таблицата Persons, за които FirstName е Иван, а LastName е Петров“. Елементарността на въпроса се състои в това, че структурираният език за заявки SQL е направен така, че правилната заявка:

`SELECT * FROM Persons WHERE FirstName = 'Иван' AND LastName = 'Петров';`³

изглежда като буквален превод на това, което се изисква във въпроса. Наличието на два въпроса като 15 и 7 не допринася за по-точното оценяване на изпитваните.

Трети по сложност е въпросът с номер 9 ($b_j = 0.0676$), който изисква да бъде определен броят на външните ключове в зададената диаграма на релационна база от данни с две таблици в отношение 1:М. Съвсем ясно е, че връзката 1:М не може да се осъществи без един такъв ключ, както и че повече от един ключ не е нужен. И все пак, понятието външен ключ не е елементарно и това прави въпроса **малко по-труден и подходящ като лек**.

Следващите по трудност, много близки един до друг, са въпроси с номера 6 ($b_j = 0.1927$) и 1 ($b_j = 0.2475$). Въпрос номер 1 изисква проследяване изпълнението на програмен фрагмент. За сложността на такива въпроси можем да съдим по два параметъра – стойностите на **колко от променливите** във фрагмента трябва да бъдат наблюдавани и **колко стъпки** на съдържащ се във фрагмента цикъл трябва да бъдат проследени. При въпрос 1 във фрагмента няма цикъл, а броят на променливите, които трябва да се наблюдават, е 2. Това прави този въпрос от най-лесните възможни от този тип, но **подходящ за горния край на интервала от лесни въпроси**. За сравнение можем да посочим въпрос номер 12 ($b_j = 0.4524$), при който се изисква наблюдаването на 12 променливи, 9 стъпки на цикъла и необичайно сравнение $A[k] < A[k-1]$, при което двете сравнявани стойности са поставени в операцията в ред, обратен на този, по който се срещат в масива. Това отчетливо определя въпроса като **добър и средно труден**.

Задачите за проследяване на код са много подходящи за тестови въпроси. Лесно може да се създаде такава задача с по-висока трудност. Например, ако във фрагмента има цикъл, който прави много стъпки и проследяването не може да се извърши стъпка по стъпка, а трябва да се намери логиката на цикъла и получаваната в края му стойност да се определи математически.

Въпрос номер 6 е подобен на въпрос номер 9, но в него става дума както за външни, така и за първични ключове и тези ключове трябва да бъдат посочени. Това поставя въпроса на горната граница на лесните и може да бъде определен като **подходящ**.

Като най-трудни от множеството средни по трудност въпроси можем да разгледаме тези с номера 3, 2 и 10. Въпрос номер 3 ($b_j = 0.6256$) касае правилата за достъп до променливи и методи на класове от обекти с различна степен на защита. Очевидно е, че това е важна област на ООП и за да се отговаря правилно на такива въпроси, се изисква както знание, така и опит. Това ги прави **подходящи** и определя трудността им в **горния край на интервала от средни въпроси**.

При въпроса с номер 2 ($b_j = 0.6824$) трябва по зададени низове да се определи в кой от четирите възможни отговора съответният израз, съдържащ метода за сравнение на низове, ще има стойност истина. Методът за лексикографско сравняване на низове е измежду най-популярните и често използвани в програмирането и въпреки това оценката на въпроса го класифицира сред най-трудните средни. Това поставя принципия въпрос семантиката на колко от многобройните методи на вградените в езиките за програмиране класове от обекти трябва да познават завършващите профила и дали в програмата за ДЗИ не трябва да се посочи едно разумно множество от методи, семантиката на които зрелостниците трябва да познават, без да се налага да ползват документацията на езика.

Въпросът с номер 10 ($b_j = 0.6934$) се оказва най-труден в темата. Той касае етапите на жизнения цикъл на един софтуерен продукт. В учебната програма на профил „Информатика“ наистина е заложено изучаването на елементи от технологиите за създаване на програмни продукти. Но характерът на обучението в профила „Информатика“ е такъв (няма производствена практика както в обичайните професионални обучения), че голямата част от учениците едва ли имат възможност да участват в създаването на сериозен програмен продукт или поне да наблюдават процеса на създаване. Затова можем да предположим, че терминологията от областта „Софтуерни технологии“ им е по-скоро абстрактна и затова въпросите от тази област се явяват трудни. Спорно е дали изобщо такива въпроси трябва да се включват в темата за ДЗИ.

4.2. За въпросите със свободен отговор

Нека сега да разгледаме оценките на въпросите с отворен отговор. Оценяването на тези въпроси с 0, 1, 2 или 3 точки е направено от оценителите в

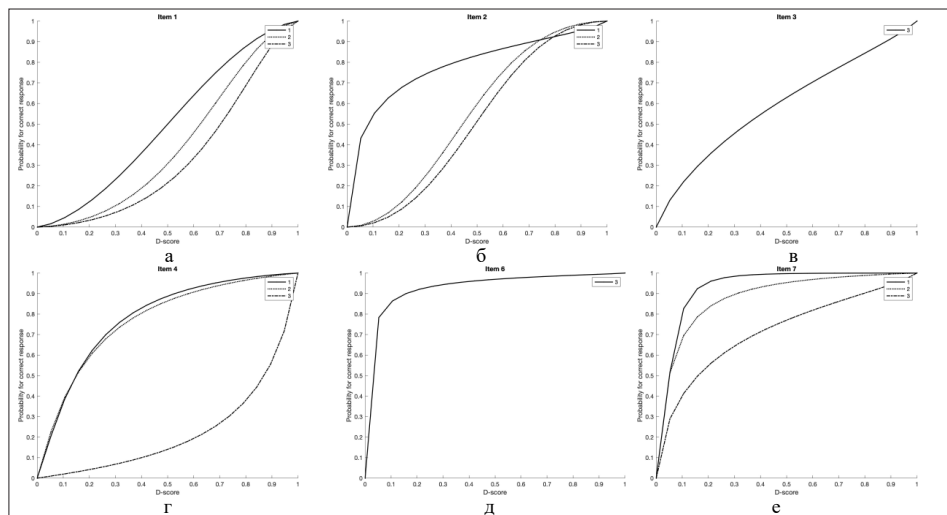
съответствие с **критерии, посочени от автора на въпроса**. Изборът на трите критерия за ненулев брой точки, както ще се види по-долу, ще се окаже неразделна част от качеството на въпроса. Оценените параметри на тестовите въпроси са представени в табл. 2.

Таблица 2. Параметри на въпросите с отворен отговор

Въпрос	Степен g	δ_j	$b_j^{(g)}$	$s_j^{(g)}$	SE $b_j^{(g)}$	SE $s_j^{(g)}$
17	1	0.5395	0.5049	1.5159	0.0269	0.2024
	2	0.6579	0.6190	1.7758	0.0245	0.2462
	3	0.7105	0.6962	1.7210	0.0255	0.2503
18	1	0.2237	0.0932	0.7653	0.0283	0.1370
	2	0.4868	0.4501	1.7762	0.0240	0.2303
	3	0.5263	0.5029	2.1083	0.0214	0.2780
19	3	0.4474	0.3583	0.8982	0.0383	0.1361
20	1	0.2368	0.1494	1.1134	0.0253	0.1656
	2	0.2500	0.1506	0.9830	0.0280	0.1517
	3	0.8289	0.8516	1.1472	0.0331	0.2056
21	1	0.2895	0.2002	1.3350	0.0244	0.1824
	2	0.3947	0.3364	1.6604	0.0243	0.2123
	3	0.5000	0.4527	1.5777	0.0260	0.2065
22	3	0.0789	0.0100	0.6528	0.0113	0.1892
23	1	0.0658	0.0356	1.1539	0.0143	0.2421
	2	0.1184	0.0475	0.9480	0.0179	0.1810
	3	0.2895	0.1668	0.7391	0.0368	0.1269
24	1	0.3026	0.2441	1.3400	0.0260	0.1792
	2	0.6053	0.5514	1.6144	0.0259	0.2180
	3	0.77632	0.7694	1.3902	0.0300	0.2185

Както и при въпросите с избираем отговор, резултатите от оценяването се визуализират отново с графиките на характеристичните криви (фиг. 5). На хоризонталната ос на графиката със стойности от 0 до 1 отново е скалата за оценката за подготвеността на ученика, а по вертикалата в този случай – вероятността за получаване на съответен брой точки при тази подготвеност. Когато данните показват наличие на отговори, оценени с 1, 2 и 3 точки, във визуализацията на оценката на един въпрос са показани три такива криви. Най-горната от трите криви показва вероятността за получаване на най-ниския брой точки, като функция на подготвеността на ученика, средната – за получаване на средния брой, а най-долната – на най-високия брой точки. Ако при оценяването не са използвани всички степени на скалата за оценка, то

показаните криви са по-малко на брой. Например, когато всички изпитвани са оценени с един и същ брой точки, кривата е само 1. Интерпретацията на характеристикните криви е същата както в случая със затворени въпроси.



Фигура 5. Статистическа оценка на въпросите с отворен отговор

На фиг. 5 са показани оценките на 6 от въпросите с отворен отговор – тези с номера 17, 18, 19, 20, 22 и 23 съответно на графики 5а до 5е. Графиките на въпроси 21 и 24 не се различават съществено от графиките на въпрос 17 и не са представени. Два от въпросите – с номера 19 (фиг. 5в) и 22 (фиг. 5д), имат оценка, състояща се от една степен. Това означава, че или въпросът не е подходящо избран, или критериите за оценяване не са достатъчно детайлни, така че да позволяват поставянето на по-ниски оценки. Друг проблем при този тип задачи може да се проследи във въпрос 20, чиито характеристикни криви са представени на фиг. 5г. При нея степените 1 и 2 на оценката са практически с еднаква трудност и не водят до по-добро разграничаване на знанията на учениците. Пример за добре балансирана е въпросът с номер 17, представен на фигура 5а и подобните на него въпроси 21 и 24.

Както се вижда от графиките на получените от анализа криви, качествата на въпросите със свободен отговор в теста доста се различават. Затова преди да ги коментираме, нека формулираме условията, които би трябвало да изпълнява един въпрос, за да бъде определен като подходящ.

1. В резултата от оценката би било добре да присъства всяка от кривите.
2. Кривите би трябвало да имат подобен и плавен растеж.
3. Не бива две от кривите (в частност и трите) да са много близо една до друга.

Както се вижда от резултатите, като **подходящи** и с **добра трудност**, **варираща от малко под средната до малко над средната** можем да определим въпрос 17 и сходните като оценка с него въпроси 22 и 24. При останалите 5 тестови въпроса има различни по характер отклонения от нормата. Затова да разгледаме по-внимателно тези 5 въпроса.

Въпрос 18 (фиг. 5б) изисква да се посочи резултатът от изпълнението на непроста рекурсивна функция (с вложено в рекурсивното извикване второ рекурсивно извикване), който е число, строго определено от дефиницията на функцията. Този въпрос очевидно не отговаря на условията. Не е ясно как, но 1 или 2 точки се получават много лесно, като вероятността да се даде отговор за 1 или 2 точки, е практически една и съща. Въпросът видимо **не е подходящ за свободен отговор** и трудно могат да бъдат посочени добри критерии за оценяване с 1 или 2 точки, когато оценяваният не е дал правилния отговор.

Малко по-различна е ситуацията с въпрос 20 (фиг. 5г). При него са зададени две таблици на релационна база от данни и заявка от тип JOIN и се пита какъв е броят на редовете и броят на колоните на получената в резултат от изпълнението на заявката таблица. При този въпрос даването на изчерпателен и точен отговор има необходимата трудност. Докато двата непълни отговора (ще предположим че 1 точка се дава за правилно посочен само единия брой, а 2 точки – само за другия) изглеждат еднакво лесни и трудно ще се намери обосновано основание кое от двете частични решения трябва да бъде оценено по-високо. Вероятно е било по-добре за тези два частични отговора да се присъжда един и същ брой точки. Ще класифицираме този въпрос, с направената забележка, като **приемлив за свободен отговор**.

Да разгледаме заедно въпросите 19 (фиг. 5в.) и 22 (фиг. 5д). Общото между тях е, че всички ученици, които са отговаряли на въпроса, са дали отговора за 3 точки, а оценките им за трудност са доста различни. Въпрос 19 според статистическата оценка е по-трудният ($b_j = 0.44737$). Той изисква да се посочи редът от програмен фрагмент, в който е допусната синтактична грешка. Тази грешка е определено груба – израз като лява част на операцията присвояване. По-високата трудност можем да отдадем на факта, че синтактически ориентираните редактори на съвременните среди за програмиране откриват синтактичните грешки още в процеса на изписване и затова учениците не се задълбочават в прецизното запомняне на синтаксиса, разчитайки, че редакторът на средата ще им напомни за направената синтактична грешка. Тук няма смислен алтернативен отговор, за който да се даде частичен брой точки, тъй като всеки такъв отговор би означавал непознаване на синтаксиса. В този смисъл, въпросът може да се определи като **добър**, но **не много подходящ за избираем отговор**.

При въпрос 22 трудността е ($b_j = 0.07894$) трябва да се проследи изпълнението на програмен фрагмент, в който има прихващане на изключения.

Тук възможностите за даване на грешни отговори са повече, но отново никой от тях не заслужава частична оценка. Затова и този въпрос **не е много подходящ за отворен отговор**. А изключително **ниската трудност** очевидно се дължи на факта, че при зададените входни данни фрагментът ще породи най-често случващото се и вероятно най-коментирано в уроците изключение `FormatException`, което няма как да не бъде добре запомнено и разпознавано.

Въпрос 23 (фиг. 5е) изисква по зададена UML диаграма на два класа да се определят кои са променливите им, тяхната степен на защитеност и връзката между двата класа. Възможностите за различни от изчерпателния и правилен отговор са много и въпросът е **подходящ** за отворен отговор. Трите криви са добре различими и в горната дясна част на чертожното поле, което ни позволява да класифицираме въпроса като **лесен**.

5. Заключение

В заключение бихме искали да отбележим, че създаването на балансирана и с голяма информативност за подготвеността на зрелостниците изпитна тема е сложен процес. Като се има предвид, за разлика от други учебни дисциплини, че разглежданият тест е първият за изучавалите профилиращия предмет „Информатика“, можем да оценим качеството му като добро. В статията сме се опитали да посочим както някои слабости, които в бъдеще да бъдат избягвани, така и реални възможности за подобряване на изпитните теми.

Много е важно тестовата част на ДЗИ да се подлага регулярно на оценки, подобни на направената тук, за да може авторите да разполагат с повече информация за качествата на предлаганите тестови въпроси и да използват тази информация за предвиждане качествата на следващите създавани теми.

Не бива да пропускаме и факта, че тестовата част носи само 40% от крайната оценка от ДЗИ, а останалите 60% се определят от резултата от практическата част, в която учениците трябва да напишат програми и заявки към релационна база от данни. Очевидно е, че качествата на задачите в тази част на ДЗИ също трябва да се подлагат на анализ. Този анализ обаче не може да бъде направен с изложената в статията методология и трябва да бъде обект на друг вид изследване. Една полезна стъпка в тази насока би била замяната на ръчното проверяване от няколко оценители с различни виждания на решението от практическата част, с формална автоматизирана проверка с множество тестови примери. Такава проверка ще намали до минимум субективността на оценяването и ще го направи по-податливо на анализ на качествата на задачите и тестовите примери.

БЕЛЕЖКИ

1. Държавен зрелостен изпит за профил „Информатика“, май 2022 г.
https://web.mon.bg/upload/30769/2DZI_INFORMATIKA_V1.pdf
2. Държавен зрелостен изпит за профил „Информатика“, август 2022 г.
https://web.mon.bg/upload/32862/2DZI_INFORMATIKA.pdf
3. В предлаганите към въпроса възможни отговори е пропуснат знакът „;“, с който трябва да завършва всяка SQL заявка.

ЛИТЕРАТУРА

- ATANASOV, D., 2020. *Delta Scoring Library for MATLAB*. <https://github.com/amitko/matlab-delta-scoring.git>
- ATANASOV, D., 2022. *D-Scoring approach*. R package. <https://github.com/amitko/DScoring.git>.
- BAKER, F.B, KIM, S.-H., 2023. *Item Response Theory, Parameter Estimation Techniques*, Second Edition. CRC Press.
- CROCKER, L., ALGINA, J., 1986. *Introduction to Classical and Modern Test Theory*. Harcourt, New York, 527.
- DIMITROV, D.M., 2020. Modeling of item response functions under the D-scoring method. *Educ. Psychol. Meas.*, 80, 126 – 144.
- DIMITROV, D.M., ATANASOV, D.V., 2021. Latent D-scoring modeling: Estimation of item and person parameters. *Educ. Psychol. Meas.*, 81, 388 – 404.
- MATLAB, 2020. version 9.8.0.1323502 (R2020a). Natick, Massachusetts: The MathWorks Inc.
- ROBITZSCH, A., 2022. *sirt: Supplementary Item Response Theory Models*. R package, version 3.12-66. <https://CRAN.R-project.org/package=sirt>.
- RAMSAY, J., WIBERG M., 2017. A Strategy for Replacing Sum Scoring. *Journal of Educational and Behavioral Statistics*, Vol. 42, No. 3, 282 – 307.
- SAMEJIMA, F., 1997. Graded Response Model. In: van der Linden, W.J., Hambleton, R.K. (eds) *Handbook of Modern Item Response Theory*. Springer, New York, NY.

THE TEST TASKS FROM THE STATE MATRICULATION EXAMINATION FOR THE INFORMATICS PROFILE DURING THE ACADEMIC YEAR 2021-2022

Abstract. In the academic year 2021/2022, the first class of Informatics Specialization in Secondary school Informatics graduated and the first State Matriculation Examinations were held for the graduates of the specialization. This article makes some quantitative and qualitative evaluations of the results of this part of the exam, containing multiple-choice and open-ended questions. The obtained results would be useful for improving the test part of the exam in preparing the topics for subsequent exams.

Keywords: profiled education; informatics; state graduating exam; statistical evaluation; multiple choice questions; open-ended questions

✉ **Dr. Dimitar Atanasov, Assoc. Prof.**

ORCID iD: 0000-0002-4079-8585

Dept. Informatics, New Bulgarian University

21, Montevideo blvd.

Sofia, Bulgaria

E-mail: datanasov@nbu.bg

✉ **Dr. Krassimir Manev**

ORCID iD: 0000-0002-0284-312X

Journal Mathematics and Informatics

125, Tsarigradsko shose blvd., bl. 5

Sofia, Bulgaria

E-mail: k.manev@azbuki.bg

✉ **Dr. Vesela Stoimenova, Assoc. Prof.**

ORCID iD: 0000-0003-2211-882X

Dept. Probability, Operations Research and Statistics

Faculty of Math and Informatics, Sofia University

5, James Bourchier blvd.

Sofia, Bulgaria

E-mail: stoimenova@fmi.uni-sofia.bg

✉ **State expert Ralica Voynova**

Ministry of Education and Science

2A, Knyaz Dondukov blvd.

Sofia, Bulgaria

E-mail: r.voinova@mon.bg