

ТЕСТОВО И ЕКСПЕРТНО ОЦЕНЯВАНЕ

Проф. Енчо Герганов,
Директор на Центъра за оценяване в
Нов български университет

Един от най-важните показатели на качеството е резултатът от обучението, изразен в числовите оценки, които преподавателите поставят при проверката и оценката на знанията, уменията и компетенциите на студентите по време на текущ контрол, изпити и други форми за установяване равнището на подготовка по съответните курсове (дисциплини). Този аспект на учебния процес винаги се е разглеждал с необходимата отговорност от страна на преподаватели, студенти и ръководители от различните равнища на управление в университета. Всички са заинтересувани оценките да отразяват по възможност най-пълно и точно знанията, уменията и компетенциите на студентите, придобити като резултат от цялостната им подготовка. В нашите университети обаче рядко се случва процесът на проверка и оценка на знанията, уменията и компетенциите на студентите да се разглежда по-задълбочено в контекста на науката за оценяване, която през последния век претърпя силно развитие както в теоретичните, така и в приложните си аспекти. Тук ще се спра само на онези въпроси, които според мене имат пряко отношение към качеството на учебния процес и решаването им би допринесло за значителното му подобрене. На първо място бих искал да посоча какво означава оценката и за какво може да се използва информацията, която тя носи.

1. Роля и функции на оценката

Оценката може да изпълнява следните функции:

- **диагностична** – показва колко успешна е работата в отделните курсове. Контрол на качеството;
- **информативна** – от оценките потребителите на кадри научават за степента на усвояване на знанията, уменията и компетенциите, които са заявени в целите на курсовете, програмите и на целия университет;
- **мотивационна** – студентът се мотивира за по-високи постижения;
- **прогностична** – въз основа на оценките могат да се правят прогнози за развитието на курсове и програми;
- **селективна** – въз основа на оценките може да се прави подбор на персонал.

2. Проблеми, свързани с установяване на резултатите от подготовката на студентите

За да се създаде работеща надеждна и валидна система за проверка и оценка на резултатите от учебния процес, трябва преди всичко да се решат следните въпроси:

- Какъв вид резултати от подготовката на студентите трябва да се установява (проверява)?
- Какви форми и методи за проверка (установяване на резултатите) могат да се използват?
- В какъв вид трябва да се представят резултатите от проверката?

Отговорът на първия въпрос като че ли е напълно ясен. Трябва да се проверява *количеството на усвоените знания, степента на формираност на уменията и равнището на придобитите компетенции*. Необходимо ли е обаче да се установява какви *нагласи* и какви *ценности* са изградени като резултат от обучението? Какви други промени у студентите очакваме да са настъпили след преживянето на пълноценен учебен процес и искаме ли да ги разкриваме?

По отношение на **втория въпрос** науката и практиката за проверка и оценка на знания, умения и компетенции предлага разнообразни форми и методи, които могат най-общо да се отнесат към две категории в зависимост от начините за приписване на числова оценка – методи, при които оценката се дава от експерт (в преобладаващите случаи **преподавател**) и методи, при които оценката се формира от броя на избраните от студента верни отговори на множество въпроси. Във втория случай участието на експертите не е пряко. Те предварително формулират множество въпроси, верните отговори към тях и дистракторите (подвеждащите отговори). Прието е създаденият по този начин инструмент да се нарича обективен тест.

Очевидно е, че в първия случай може да се използва голямо разнообразие от форми, при които се проявява подготовката на студентите, като *писмена работа по зададена тема, есе, експериментална разработка, решаване на задачи, анализ на проблем, курсов проект* и др. Студентските разработки се проверяват и оценяват от експерти (преподаватели).

Във връзка с третия въпрос според установената у нас практика резултатите от проверката се представят във вид на числова оценка – една от оценките на шестобалната система, която е регламентирана чрез Закона за висшето образование и Закона за народната просвета.

Когато преподавателите оценяват писмените работи на студентите, поставят оценка от шестобалната система. Когато пък изпитването е чрез тест, обикновено се отчита броят на тестовите задачи, за които е даден верен отговор. След това скалата се трансформира в шестобална.

Това не означава обаче, че не можем да използваме и други числови скали, които бихме получили, като използваме методи, разработени в теорията на тестовете (когато прилагаме обективен тест) и теорията на скалирането (когато прилагаме експертно оценяване).

В следващите раздели ще разгледам тези въпроси малко по-подробно.

3. Аспекти на знанията, уменията и компетенциите и подходите за тяхната проверка

3.1. Видове прояви на знанията

Както е известно от теорията и практиката на оценяването, има различни прояви на усвоените знания и различни подходи за проверката им, като репродуктивност, продуктивност и творческо прилагане на знанията, уменията и компетенциите.

Репродуктивното знание се проявява, когато се:

- изброяват факти;
- дефинират понятия;
- изразяват принципи
- формулират закони;
- възпроизвеждат формули и др.

Проявите на **продуктивното знание** пък са следните:

- осмисляне на фактите;
- опериране с понятията;
- разбиране на принципите;
- прилагане на законите;
- умение да се прилагат пряко отделни знания за решаване на сравнително прости задачи, проблеми.

По-висша степен на усвояване на знанията са техните **приложения, преноса им** при решаване на различни задачи. Към тази категория можем да отнесем:

- уменията за анализ и синтез;

- оценяването на ситуации и проблеми;
- защита на цялостна концепция, теза;
- умение да се прилагат знания от различни области с цел да се реши сложен проблем.

Най-висшата степен на усвояване на знанията се проявява в **творческото им прилагане**, което се изразява в:

- умения да се проблематизира;
- правене на изводи и предположения при недостиг на информация;
- подбор на адекватни методи за изследване;
- нетрадиционна интерпретация на резултати;
- избор на нетрадиционни подходи към решаване на проблеми и разкриване на противоречия;
- предлагане на алтернативни решения;
- преценяване на риск.

3.2. Форми и методи за проверка на видовете прояви на знания, умения и компетенции

Тук ще се опитам да дам някои от най-подходящите форми за проверка и оценка на знания при различните равнища на техните прояви.

В практиката на изпитване в университетите и в училище за съжаление най-често обект на проверка и оценка е **репродуктивността на знанията**. Това се постига чрез:

- обективен тест;
- писмено или устно възпроизвеждане на знания в дадена област и др.

Продуктивността на знанията може да се провери чрез:

- решаване на прости задачи и казуси;
- есе;
- рецензия;
- критика;
- участие в дискусии и др.

Подходящи форми за проверка на **приложението, преносът на знания** са:

- изследователски разработки;
- доклади;
- есета;
- решаване на сложни задачи и казуси, проекти;
- създаване на модел...

Най-висшата степен на владение на знанията - **творческото им прилагане**, може да се установява чрез следните форми:

- планиране и провеждане на оригинално изследване;
- анализ на актуални проблеми;
- анализ на специализирана литература;
- участие в проблемни дискусии.

4. Подходи към оценяване на продуктите от проверката на знанията, уменията и компетенциите

Когато се разработва процедура за проверка и оценка на знания, умения и компетенции, трябва да се реши какъв подход да се използва. Според теорията на тестовете съществуват два подхода към измерването на знанията – **критериен** и **нормативен**.

4.1. Критериен подход

Критерийният подход изисква постиженията на студента да се сравняват с предварително зададен критерий.

В съответствие с критерийния подход трябва:

- да се направи списък на всички предвидени за усвояване знания (факти, понятия, принципи, закони и пр.) в курса;
- да се установи чрез някои от методите и формите за проверка какъв процент знания от този списък студентът е усвоил на всяко от четирите равнища – репродуктивност, продуктивност, пренос и творческо приложение;
- едно и също ключово знание да се проверява няколко пъти в различни контексти

При критерийното оценяване напълно е възможно всички студенти да получат отлични оценки (когато са усвоили 100% от предвидения материал) или слаби оценки (когато процентът на усвояването е нулев).

4.2. Нормативен подход

Твърдението, че не може всички студенти да имат *отличен*, отразява нормативен подход към оценяването. То съдържа допускането, че студентите се различават по способности, творчески възможности и др. Ако всички се оценяват с най-високата оценка, се пренебрегват посочените различия. Според критерийния подход обаче тези различия не са обект на оценяване.

В съответствие с нормативния подход трябва:

- да се определи нормативната група, спрямо която ще се сравняват постиженията на всички студенти;
- да се използват стандартизирани оценъчни процедури – стандартизирани тестове, калибрирани оценки на проверители и др.

Както се вижда, при нормативния подход постиженията на един студент се сравняват с постиженията на останалите студенти в групата. При това сравнение е естествено да се отразят индивидуалните различия между тях. Когато се използва този подход, възниква проблем как да се зададе **групата-норма**, спрямо която ще се сравняват резултатите на всеки студент. В различни години и в различни групи може да има различни средни равнища на постижения.

5. Бални системи и скали, които се използват при оценяването

В практиката на оценяване най-често се използват две групи скали:

- скали, които се използват при субективното оценяване (оценяване от експерти);
- скали, които се използват при обективните тестове

5.1. Оценяване от експерти

Както вече посочих, съществуват множество форми на проверка, които водят до продукти, изискващи оценяване от експерти. Ето някои от методите, които се предлагат в теорията на скалирането:

- методът на равноизглеждащите интервали, т.н. Ликертови скали. Типичен пример – българската шестобална система. Тук попадат 10-балните, 20-балните, 100-балните и всякакви други N-бални скали. У нас се използва предимно този метод във варианта шестобална скала;
- методът на ранжирането. Всички писмени разработки се ранжират от най-добрата към най-лошата. Така оценките са във вид на рангове;
- Методът на сравняване по двойки. Всяка разработка се комбинира с всяка друга. За всяка двойка експертът трябва да каже коя разработка е по-добра. Скалата се получава чрез прилагането на специални методи за анализ на резултатите от сравняването.

5.2. Проблеми при експертното оценяване

Изследванията показват, че не всеки специалист в областта си може да бъде едновременно и добър експерт-оценител. Добрият експерт-оценител удовлетворява следните критерии:

- оценките му се съгласуват във висока степен с оценките на множество други проверители, които оценяват едни и същи писмени работи;
- рядко попада в екстремните позиции на континуума *строгост-снизходителност*;
- лесно се научава да оценява писмени работи по зададени изисквания.

Тези критерии могат лесно да се операционализират.

5.2.1. Критерият *съгласуваност* на оценките

Същността на този критерий може да се покаже чрез един пример. Нека да вземем пет проверители, които са поставили оценки на шест писмени работи (вж. табл. 1)

Таблица 1: Оценки, които пет проверители са поставили на шест писмени работи

ПИСМЕН И РАБОТИ	ПРОВЕРИТЕЛИ				
	А	Б	В	Г	Д
1	4,00	5,50	5,00	6,00	4,00
2	3,75	5,25	4,50	5,50	3,25
3	3,50	5,00	4,75	5,00	5,00
4	3,25	4,50	4,25	5,25	3,50
5	3,00	4,75	4,00	4,75	4,75
6	2,00	4,75	3,75	4,50	4,50
Средна оценка:	3,25	4,96	4,38	5,17	4,17

Дори с просто око може да се види, че оценките на първите четирима проверители се съгласуват във висока степен. Ранжирането на писмените работи въз основа на оценките на тези проверители от най-високата до най-ниската оценка е почти едно и също, а именно **А**:1,2,3,4,5,6; **Б**:1,2,3,5,6,4; **В**:1,3,2,4,5,6; **Г**:1,2,4,3,5,6. Ранжирането по оценките на петия проверител **Д** обаче силно се различава от първите пет ранжировки, а именно **Д**: 3,5,6,1,4,2.

Степента на съгласуваност между оценките се измерва с коефициента на корелация. Коефициентите на корелация между оценките на всеки двама проверители от горния пример са дадени в табл. 2.

Таблица 2: Коефициенти на корелация между оценките на пет проверители за шест писмени работи, дадени в табл. 1

	А	Б	В	Г	Д
А	1,00	0,91*	0,91*	0,88*	-0,38
Б	0,91*	1,00	0,89*	0,84*	-0,23
В	0,91*	0,89*	1,00	0,84*	-0,15
Г	0,88*	0,84*	0,84*	1,00	-0,58
Д	-0,38	-0,23	-0,15	-0,58	1,00

* $p < 0,05$

Както се вижда от табл. 2, коефициентите на корелация между всеки двама от първите четирима проверители са много високи (варират от 0,84 до 0,91) и са статистически значими, докато корелациите между оценките на петия проверител с оценките на всеки един от останалите четирима са много ниски и статистически несъществени. Това показва, че петият проверител или не може да оценява, или оценява нещо, което никой от останалите не отчита.

5.2.2. Критерият *строгост/снижодителност*

Ако разгледаме внимателно данните на първите четирима проверители в табл. 1, ще видим, че техните оценки за една и съща работа се разминават съществено. Например за писмена работа № 1 са получени оценките 4; 5,5; 5; 6; 4, за писмена работа № 2 – 3,75; 5,25; 4,5; 5,5; 3,25 и т.н. Какво тогава им е качеството на тези четирима експерти-оценители, когато оценките им толкова съществено се разминават.

Тук се е проявило съществено свойство на оценителите, което трудно може да се промени в желана посока чрез обучение. Това е свойството *строгост/снижодителност* на експерта-оценител. По данните в табл. 1 може да се каже, че най-строг е проверител А (средната оценка по оценките му на шестте писмени работи е 3,25), а най-снижодителен – проверител Г (неговата средна оценка пък е 5,17). При конкурс за подбор на проверители и липса на процедура за калибриране на оценките, получени при конкурсен изпит, споменатите двама проверители и други подобни на тях не би трябвало да се включват в екипа поради прекалената им снижодителност или строгост.

5.2.3. Критерият *обучаемост*

За да се установи в каква степен експертите-оценители са податливи на обучение, кандидатите за оценители се подлагат на оценяване по посочените по-горе критерии ПРЕДИ и СЛЕД обучението в системата от изисквания за оценка на писмени работи по конкретна тема. Степента на промяна в посока на по-голяма степен на съгласуване на оценките на кандидата с оценките на другите проверители, както и на отклоняването му от екстремните точки на строгостта и снижодителността след обучението ще бъде доказателство за обучаемостта му.

Разгледаните особености в оценъчното поведение на експертите-оценители се отнасят за всички университетски преподаватели, които оценяват писмени работи или устни отговори на студентите. Те трябва да се вземат предвид, когато се разработва система за качествено оценяване на постиженията на студентите.

5.3. Оценяване чрез обективни тестове

Както вече казах, обективният тест е съвкупност от тестови задачи, всяка от които дава 1 т. (когато е избран верният отговор към нея) или 0 т. (когато е избран някои от подвеждащите отговори или не е посочен никакъв отговор).

Резултатите от тестовата проверка на знанията се представят чрез няколко вида скали:

- *суров бал* – брой на правилните отговори по всички тестови задачи;
- *процентилна скала* – постижението на студента се измерва с процента на тестираните, които имат по-нисък суров бал от неговия;
- *стандартни скали* – суровият бал се трансформира в Z-единици (единици на нормираното нормално разпределение);
- *скалови стойности*, получени чрез прилагане на *логистични модели*.

По-нататък ще се спра по-подробно на някои основни понятия от теорията на тестовете, за да дам представа какво трябва да знае преподавателят, когато иска да използва обективен тест за измерване постиженията на своите студенти.

6. Някои основни понятия от теорията на тестовете

В теорията на тестовете се постулира наличието на континуум на знанията, уменията и компетенциите. Тестовите задачи и студентите са представени като наредени точки на този континуум – студентите според количеството знания, умения и компетенции, които притежават, а задачите според трудността им или другояче казано според количеството знания, умения и компетенции, които са необходими, за да бъдат решени.

Теорията изисква тестовите задачи да са в такъв формат, че да дават 1 т., когато се посочи верен отговор, и 0 т., когато се даде някакъв друг отговор (стандартен формат). Колкото студентът е по-надаясно на споменатия континуум, толкова повече задачи ще реши, т.е. толкова по-голям брой точки ще получи.

Ето един пример на тестова задача по физика със стандартен формат:

Ако увеличим разстоянието между два точкови заряда два пъти, електричната сила на взаимодействие между тях:

- а) се увеличава два пъти;
- *б) намалява четири пъти;
- в) намалява два пъти;
- г) се увеличава четири пъти;

(Верният отговор е отбелязан със звездичка.)

За съжаление понякога преподаватели, които искат да прилагат тестови процедури за измерване на постиженията, използват нестандартни формати, за които няма разработена психометрична теория. Ето един пример на задача по биология с нестандартен формат.

Изброени са съединения. Някои от тях са свързани с нуклеиновите киселини.

Изградете молекулите на ДНК и РНК, като поставите съответните цифри на обозначените с многоточие места:

- | | | |
|-------------------|----------------------|--------------|
| 1. дезоксирибоза; | 6. фосфорна киселина | 11. глицин; |
| 2. рибоза; | 7. тимин; | 12. цистеин |
| 3. глюкоза; | 8. аланин; | 13. цитозин |
| 4. холин; | 9. урацил; | 14. лизин; |
| 5. аденин; | 10. гуанин; | 15. Хистидин |

ДНК..... РНК.....

Отговор:

1,5,6,7,10,13

(12x0,2=2,4 т.)

2,5,6,9,10,13

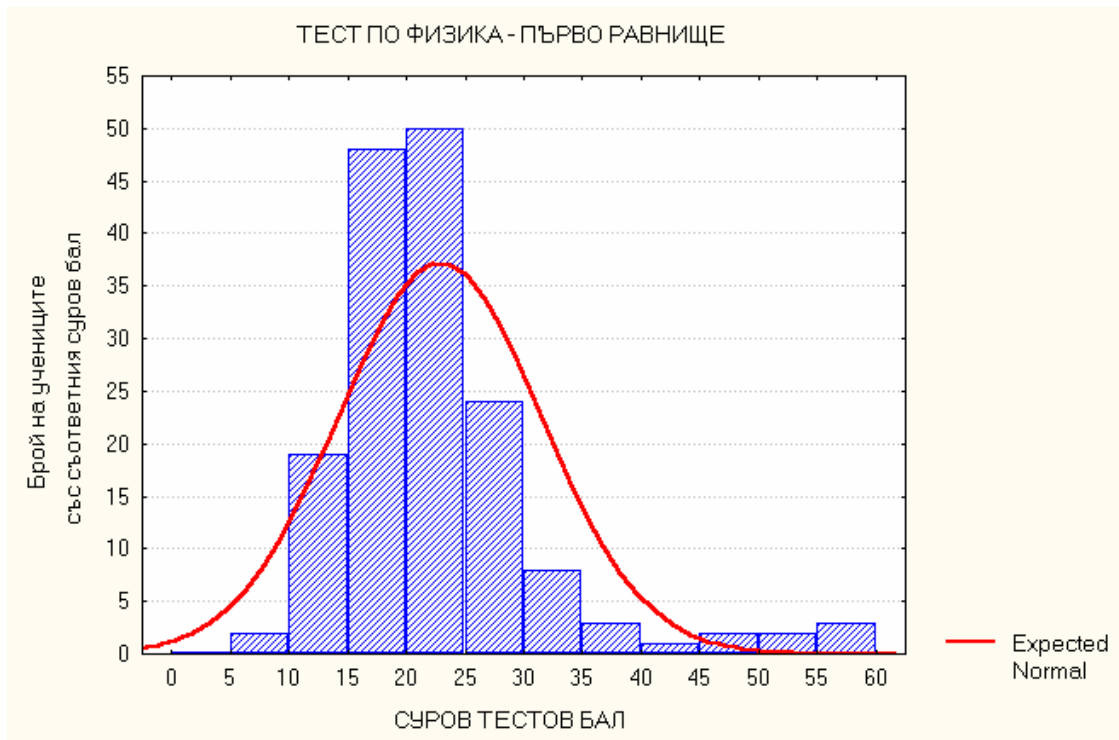
Системата за точкуване тук води до бална скала, която трудно се обхваща от основните постановки на класическата теория на тестовете. Оттук нататък ще разглеждам само тестови задачи със стандартен формат и тестове, съдържащи такива задачи.

6.1. Разпределения на суровия тестов бал

Броят на вярно решените задачи със стандартен формат (1,0) образува суровия бал на студента. За да се изследват някои важни характеристики на теста, е необходимо да се представи честотното разпределение на суровия бал на студентите от дадена група. По характера на това разпределение може да се съди за трудността на теста.

ТРУДЕН ТЕСТ

Когато кривата на разпределение е асиметрична и скосена отляво, тестът е труден. На фиг. 1 е показана хистограмата и кривата на разпределение на тест по физика, който се е оказал много труден за групата. Под фигурата са дадени статистиките на това разпределение. Примерите са от пилотната матура през 2002 г.



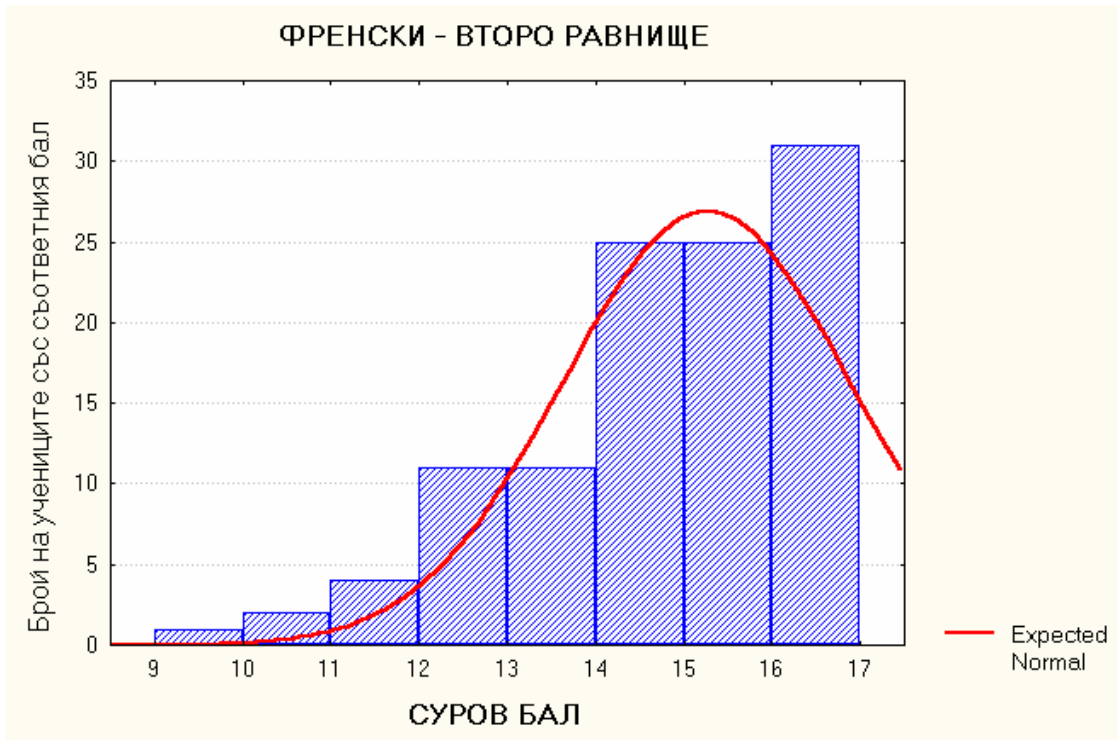
**ФИГУРА 1. ЧЕСТОТНО РАЗПРЕДЕЛЕНИЕ НА СУРОВИЯ БАЛ
в теста по физика – първо равнище** **СРЕДЕН БАЛ: 22,9**

МЕДИАНА: 21,5
ДОЛЕН КВАРТИЛ: 17,0
ГОРЕН КВАРТИЛ: 26,0
АСИМЕТРИЯ: 1,9
ЕКСЦЕС: 5,1

Както се вижда, коефициентът на асиметрия е положителен и много голям, което говори за труден тест. С други думи по-голямата част от тестираните ученици са струпани около ниските стойности на сурови бал.

ЛЕСЕН ТЕСТ

На фигура 2. е дадена хистограмата и кривата на разпределение на лесен тест с отрицателна асиметрия, т.е. Кривата е скосена отдясно. По-голямата част от тестираните са струпани в зоната на високия суров тестов бал.

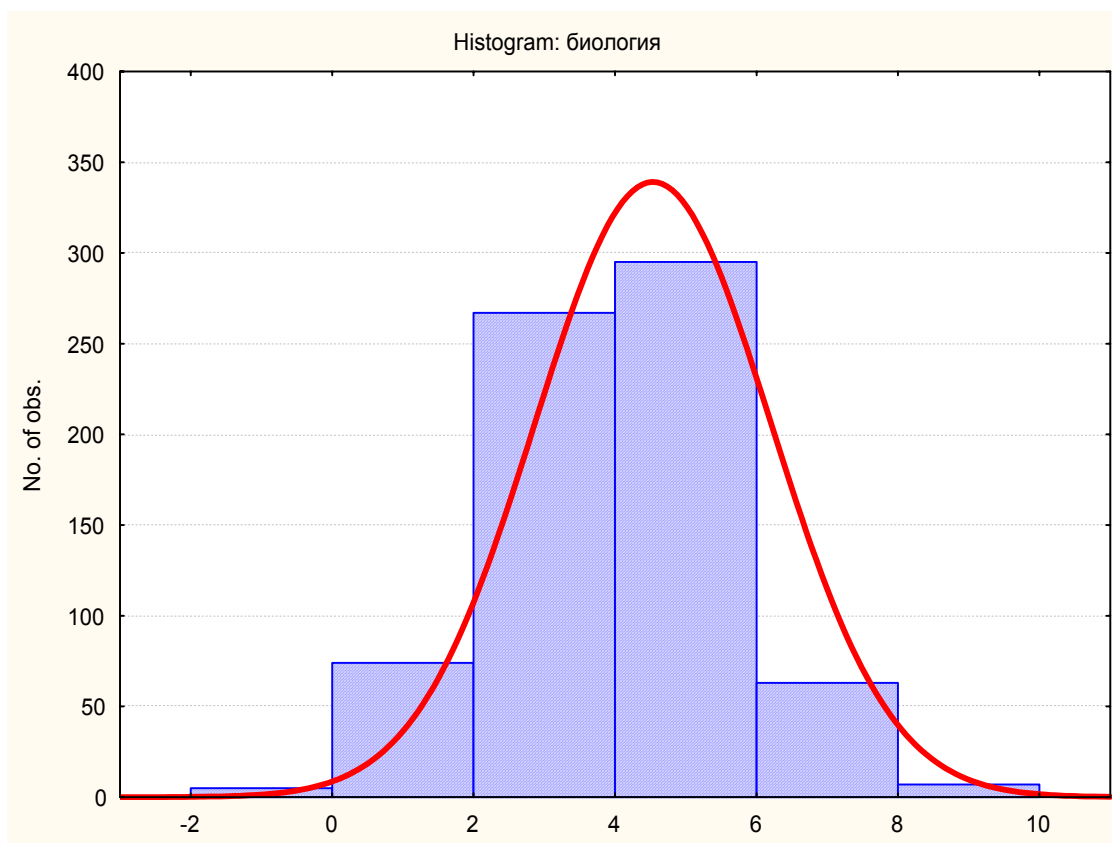


ФИГУРА 2. ФРЕНСКИ ЕЗИК – ВТОРО РАВНИЩЕ

Статистика	Стойност
<i>Средна</i>	15,26
<i>Медиана</i>	16.00
<i>Минимум</i>	10.00
<i>Максимум</i>	17.00
<i>Долен квантил</i>	14.00
<i>Горен квантил</i>	17.00
<i>Станд. откл.</i>	1.63
<i>Асиметрия</i>	-0.90
<i>Ексцес</i>	0.34

ТЕСТ С НОРМАЛНА ТРУДНОСТ

На фиг. 3 е дадена хистограмата и кривата на разпределение на субтеста по биология от Тест за общообразователна подготовка за приеман изпит в НБУ. Субтестът е с нормална трудност. Вижда се, че коефициентът на асиметрия е много малък (-0,009)



ФИГУРА 3. ЧЕСТОТНО РАЗПРЕДЕЛЕНИЕ НА СУБТЕСТ ПО БИОЛОГИЯ ОТ ТОП

Статистика	Стойност
<i>Средна</i>	4,53
<i>Медиана</i>	5,0
<i>Минимум</i>	0
<i>Максимум</i>	10
<i>Долен квартил</i>	3
<i>Горен квартил</i>	6
<i>Станд. откл.</i>	1.67
<i>Асиметрия</i>	-0.009
<i>Ексцес</i>	0.01

6.2. Анализ на тестовите задачи

Качеството на тестовите задачи се оценява по четири нейни характеристики – трудност, различителна сила, ефективност на дистракторите и съгласуваност с общия тестов бал.

Трудността се измерва чрез процента на тестираните, които са решили задачата.

Различителната сила се измерва по следния начин. Всички тестирани се ранжират по намаляване на суровия бал – от студенти с най-висок суров бал към студентите с най-нисък бал. Взимат се първите 25% студенти, които образуват силната група, и последните 25%, които образуват слабата група. След това се наблюдава поведението на задачата в двете групи. Ако

всички студенти от силната група са решили задачата и нито един от слабата не я е решил, задачата има максимална различителна сила. Ако процентът на решените задачи от силната група е равен на процента на решените задачи от слабата група, тази задача е с нулева различителна сила.

Съгласуваност с общия бал. Измерва се чрез коефициента на корелация между отговорите на задачата и суровия тестов бал.

Ето някои примери за добри и лоши задачи според тези критерии.

Пример за **добра** задача.

Напрежението между т. *A* и т. *B* е 3 V . Работата, извършвана от електрическата сила, при преместването на заряд $2\text{ }\mu\text{C}$ от т. *A* в т. *B*, е:

1. $6\text{ }\mu\text{J};*$
2. $-6\text{ MJ};$
3. $1,5\text{ MJ};$
4. $-1,5\text{ }\mu\text{J}.$

Трудност (Процент решили задачата)	Различителна сила	Корелация с общия бал
0,64	0,75	0,63

Ефективност на дистракторите

Отговори	Относителна честота на отговорите в цялата група	Относителна честота на отговорите в слабата група	Относителна честота на отговорите в силната група
1*	0,64	0,25	1,0
2	0	0	0
3	0,29	0,5	0
4	0	0	0
други	0.07	0	0

Пример за **лоша** задача.

Без йонизация **не е възможно** протичането на електричен ток в:

1. метали;
2. електролити;
3. газове;*
4. полупроводници.

Процент решили задачата	Различителна сила	Корелация с общия бал
0,50	-0,35	-0,23

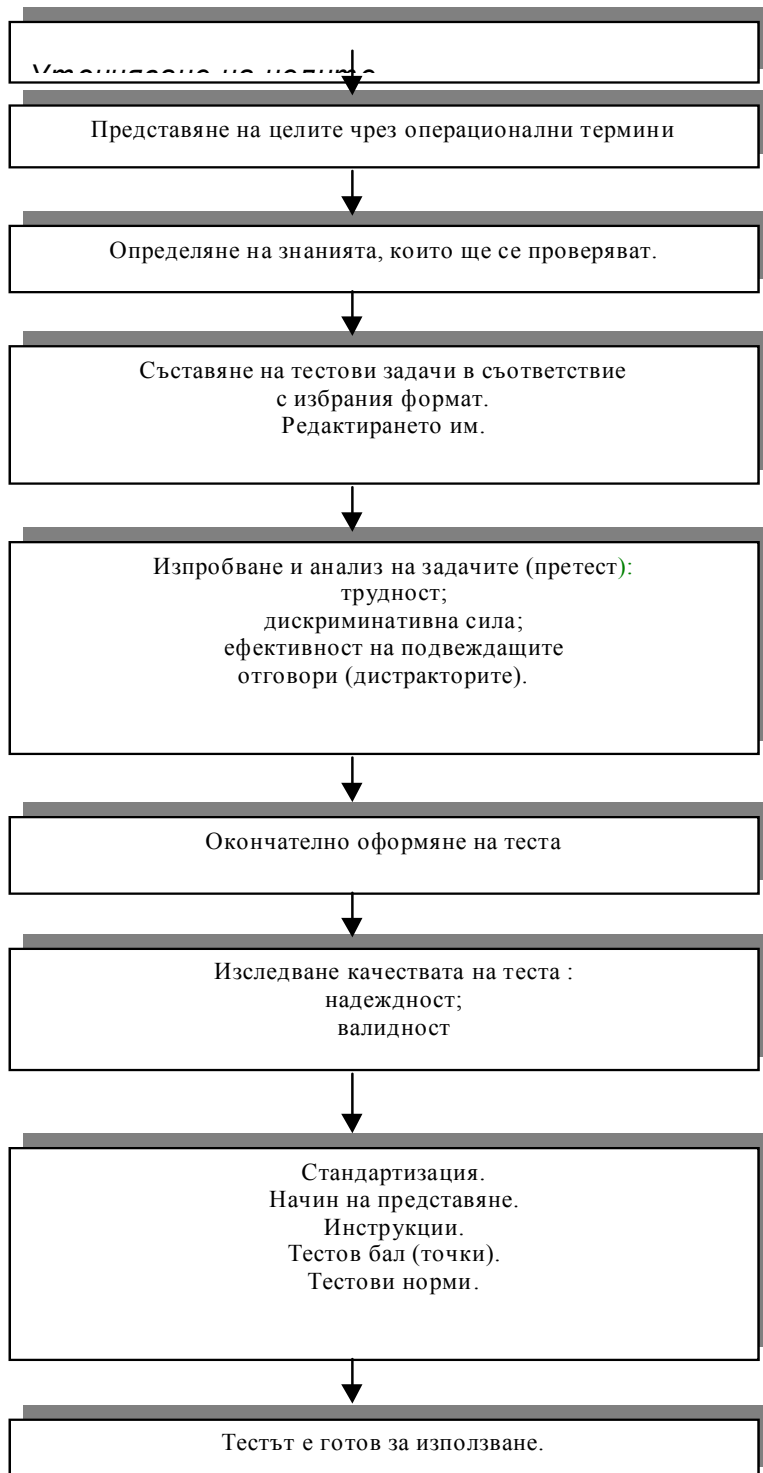
Отговор	Относителна честота на отговорите в цялата група	Относителна честота на отговорите в слабата група	Относителна честота на отговорите в силната група
1	0,21	0	0
2	0,21	0,25	0,40
3*	0,50	0,75	0,40
4	0,07	0	0,20
други	0	0	0

В последната задача се е получила отрицателна стойност на различителната сила, което означава, че по-голям процент от учениците в слабата група са решили задачата в сравнение с процента на учениците от силната група, които са се справили с нея. Това е парадоксално и обяснението е, че задачата е лоша и трябва да се бракува.

6.3. Етапи, през които се преминава, за да се конструира качествен обективен тест

Разработването на обективен тест изисква да се изпълнят някои важни изисквания, които са представени в долната схема.

**КОНСТРУИРАНЕТО НА ТЕСТ ЗА ИЗМЕРВАНЕ НА ЗНАНИЯТА, УМЕНИЯТА И КОМПЕТЕНЦИИТЕ ПРЕМИНАВА
ПРЕЗ ЕТАПИТЕ, ПОКАЗАНИ НА СХЕМАТА**



7. Заключение

В доклада си само поставих някои проблеми, свързани с надеждното и валидното оценяване на знанията, уменията и компетенциите на студентите. Желанието за промяна в този важен за университета процес би трябвало да предизвика система от квалификация на преподавателите в областта на теорията и практиката на оценяването, като и разработване на информационна система за анализ на резултатите от тестиранията и непрекъснато усъвършенстване на оценъчните инструменти.