

*Иновации, предизвикателства и тенденции в постмодерното образование
Innovations, Challenges and Tendencies in the Post Modern Education*

СТАТИСТИЧЕСКИ АНАЛИЗ НА ИЗПИТИТЕ ПО ОБЩА ЕЗИКОВА КУЛТУРА В ТРАКИЙСКИ УНИВЕРСИТЕТ

Николай Такучев
Тракийски университет

Резюме. Тази статия разглежда проблемите от доцимологичен характер, свързани с евристичната оценка на проведените писмени изпити по Обща езикова култура за Тракийски университет, резултатите от които се използват от Университета за допускане и класиране на кандидатите. Обсъждат се възможни подходи за намаляване на отрицателното въздействие на грешна оценка при класирането на кандидати с подобни (близки) оценки поради важните жизнени последици на оценката за учениците и за целите на приема в Университета. В тази връзка се предлага 33-степенна скала, чиито класове имат обхват, пропорционален на разпределението на грешката в обхвата на оценките. Използването на подобна скала, чиито класове зависят от грешката на оценката, прави процеса на оценяване точен, което намалява риска от неправилно класифициране.

Keywords: docimology, measurement in education, scale, grade, error of the educational assessment, uncertainty of the evaluation

Увод

Доцимологията (от гръцки – тестология)¹⁾ е дял от дидактиката, чийто обект на изследване най-общо е оценяването на знания, разглеждано както като процес на оценяване, така и като резултати от оценяването.

Кандидат-студентският прием е по същността си доцимологичен процес, на първия етап от който се оценяват кандидатстудентските знания, след което на втория етап кандидатстудентите се класират по специалности в зависимост от оценките си на първия етап.

Цел на настоящата работа е да изложи проблемите от доцимологичен характер, свързани с оценяването на писмените работи от изпитите по Обща езикова култура за Университета.

Материал и методи

Данни

За анализа са използвани оценките, получени в резултат на работата на оценителските комисии за 15 кандидатстудентски изпита по Обща езикова

култура – предварителни и редовни, в Тракийски университет за последните 5 години – от 2009 г. до 2013 г. (плюс един от 2008 г.). Всяка година се провеждат по два предварителни и един редовен изпит по Обща езикова култура за специалностите на Университета.

Оценяващи комисии

За дисциплините с писмени кандидатстудентски изпити всяка година се формират комисии от оценители на писмените работи. Комисиите, оценяващи писмените работи от изпитите по Обща езикова култура през изследвания петгодишен интервал, остават с почти неизменен състав – броят на оценителите е 4 или 5, към основното неизменно през годините ядро от членове е добавян или отпада по един оценител. За петгодишния интервал са оценени общо 1299 кандидат-студенти. Оценените писмени работи са повече – 2356, тъй като част от кандидат-студентите са се явявали на изпит повече от един път.

Оценяване на конкурсните работи

Оценяването на кандидатстудентски знания е особено отговорно, тъй като въз основа на оценките си кандидат-студентите се класират или не – т.е. това оценяване има съдбовно влияние върху бъдещия живот на младите хора.

При оценяването на знания в конкурсите в България традиционно се използва числена оценка в интервала между 2 до 6. Интервалът на оценките може да се онагледя графично като отсечка с крайни точки 2 и 6, в която на всяка точка съответства оценка. Съгласно тази представа оценките са съвкупността от точки (по-долу наричани точкови оценки), формиращи интервала на оценките. Не всички оценки участват в класирането. В конкурсите има заложен критерий за допуск до участие в класирането – минимална оценка, по-висока от долната граница на интервала на оценките, оценките над която участват в класирането, а тези под нея – не. Обикновено оценката за допуск до класиране е 3,00.

При евристичното оценяване на знания оценителят експерт оценява знанията, изложени в анонимната писмена работа, съпоставяйки им числена оценка съобразно експертните си знания и критерии. Писмената работа се състои от няколко части, които оценителят оценява поотделно и намира средноаритметичната им стойност, която е закръглена до стотна. Тя е окончателната оценка на експерта за знанието, показано в писмената работа.

Крайната точкова оценка за писмената работа е средноаритметичната от оценките на двама независими оценители и се изчислява също с точност до стотната.

Експертна грешка при оценяването

Оценката винаги е натоварена с неточност, т.е. получената оценка се отличава от „точната“, която би поставил „идеалният“ оценител.

Разликата между дадената от експерта оценка и съответстващата на оценяваното знание точна оценка е мерило за неопределеността на оценката вследствие на грешката на оценителя. Крайната оценка на писмената работа от двамата оценители е натоварена с грешката на двамата оценители в двойката. Ако е из-

вестна съвкупността от грешките при оценяването на писмените работи, може да се определи средната грешка на оценителската комисия като цяло.

Неопределеността в оценките прави некоректно сравняването на близки оценки. Когато оценяването е за целите на класиране, в което много кандидати се борят за малко места, е коректно подреждането на кандидатите по оценките за знанията им да бъде съобразено с грешката на оценяване на комисията. Пример: ако двама кандидати са с оценки съответно 4,40 и 4,60, разликата в оценките на двамата е 0,20 при средна от двете оценки 4,50. Ако грешката на комисията е 0,50, то не може да се твърди, че знанията на втория кандидат са повече от тези на първия, тъй като при такава грешка точната оценка на втория кандидат-студент е някъде в интервала от 4,00 до 5,00 (от $4,50 - 0,50 = 4,00$ до $4,50 + 0,50 = 5,00$). При класиране на кандидат-студентите по оценки поради грешката на оценителската комисия съществува риск кандидат с повече знания да бъде изместен в класирането от такъв с по-малко знания поради грешката при оценяването.

Рискът от некоректно класиране е свързан, от една страна, с неопределеността на оценката, дължаща се на грешките при оценяването, и от друга – с гъстотата на оценките в даден подинтервал от интервала на оценките. Колкото е по-голяма грешката на оценяване и колкото по-близки са оценките на конкуриращите се кандидат-студенти, толкова е по-голям рискът от неправилно класиране.

Множество мерки целят намаляване на риска от (умишлена) оценителска грешка:

- забранява се участие в комисии за оценяване на близки на кандидатите или на оценители, водещи кандидат-студентски курсове;

- конкурсните работи са шифрирани, т.е. анонимни на етапа на оценяването – първо се пише оценката и после се дешифрира работата, за да се идентифицира кандидат-студентът, чиято е писмената работа;

- всяка конкурсна работа се оценява независимо минимум от двама оценители, работата на оценяващите комисии е организирана така, че оценителите да не знаят с кого от колегите си оценяват дадена работа;

- ако разликата в оценките превиши определена стойност, тя се оценява от трети оценител.

Тези мерки обаче не намаляват случайната компонента на оценителската грешка. По-долу се обсъждат мерки за намаляване на влиянието на случайната компонента на експертната грешка върху класирането.

Грешка на оценяването на писмена работа и средна грешка на оценителските комисии

По данните от работата на изследваните комисии бяха изчислени грешките при оценяване на писмените работи – както индивидуалните, така и на комисията като цяло. Като характеристика на грешката на оценяване на писмената работа беше използвано средното квадратично отклонение, изчислено от оценките на двамата независими оценители на една и съща писмена работа.

Пресмятането показва, че в случая то е *половината от разликата между двете независими оценки*. Средноаритметичната от грешките за отделните писмени работи е използвана като характеристика на грешката на комисията като цяло.

Ска̀ла от балове

Поради наличието на грешка на комисията може да се говори за зони на неразличимост на близките оценки. Тези зони са по-големи в участъците от интервала на оценките с по-голяма средна грешка на комисията. Негативният ефект от неразличимостта върху коректността на класирането се увеличава в участъците с голяма гъстота на оценките.

Негативният ефект може да бъде намален, ако:

- се прилагат подходящи критерии за оценяване, разпределящи колкото е възможно по-равномерно оценките в интервала от 2 до 6;

- на група близки оценки бъде съпоставен еднакъв бал, т.е. се въведе скала от балове, покриваща интервала на оценките от 2 до 6. Скалата е съвкупност от незастъпващи се балове – подинтервали на интервала на оценките, покриващи целия интервал на оценките. Всички оценки в рамките на даден бал се смятат за неразличими, с еднакъв бал, например „шестобалната“ скала съдържа баловете „слаб“, „среден“, „добър“, „много добър“ и „отличен“. Баловете имат:

- а) *обхват* – част от интервала на оценките. Например обхватът на бал „отличен“ е подинтервалът от 5,50 до 6,00;

- б) *големина* – показва поредността на бала в скалата. Например бал „добър“ е по-голям от бал „среден“ и по-малък от бал „много добър“, използва се скала от балове, пропорционални по обхват на грешката на комисията. Най-често при евристичното оценяване се използва скала с предварително фиксирани балове, несъобразени по обхват с грешката на оценяващата комисия. Под разпределение на грешката на оценяващата комисия в рамките на интервала на оценките от 2 до 6 по-долу се разбира зависимостта на средната грешка на комисията от средната оценка. Ако грешката на комисията е еднаква за целия интервал на оценките, то споменатата пропорционалност между обхвата на баловете и грешката на комисията е изпълнена за всяка скала. На практика най-често разпределението на грешката е по-сложна функция на оценката, т.е. има стойности на оценката, около които комисията оценява по-неточно, с по-голяма грешка от тази около други стойности на оценката. В такъв случай за оптимално оценяване на знанията с цел последващото им сравняване (класиране) е целесъобразно използването на скала с балове, чийто обхват е пропорционален на грешката на комисията в подинтервала на оценките в обхвата на бала.

Броят на баловете в скалата може да бъде избран от допълнителни съображения. Например при малък брой балове в скалата множеството кандидати с еднакъв бал биха затруднили класирането при ограничен брой на местата поради повличане в бала – наличие на повече кандидати с еднакъв бал от-

колкото места за специалността. Този проблем може да се реши, като броят на баловите се направи достатъчно голям и същевременно се запази съотношението в обхватите им, пропорционално на грешката на оценяващата комисия. Проблемът с повличането намалява и чрез използване при класирането на математически израз (също наречен „състезателен бал“), включващ както оценки от изпити, така и такива от дипломата.

В настоящата работа е предложена бална скала с обхвати на баловите, пропорционални на средната грешка на комисиите по оценяване на писмените работи от изпита по Обща езикова култура. Броят на баловите може да бъде подчинен на различни критерии. За споменатата скала броят на баловите беше избран така, че средната големина на бала в скалата да е равна на максимума на средната грешка на комисиите.

За разработването на скалата беше получен регресионен степенен модел на зависимостта на средната грешка на комисиите от оценката. Тази зависимост е наричана по-долу *функция на грешката*. Данни за определянето на модела бяха стойностите на средната грешка на комисиите, изчислена като средна стойност от грешките за оценките на отделните писмени работи в двадесет подинтервала с обхват 0,2 – от 2 до 2,19, от 2,2 до 2,39 и т.н., до 6. Моделът е определен като степенен тренд чрез съответната вградена функция на MS Excel.

Функцията на грешката на комисиите е характеристика на неточността (неувереността) на членовете на комисията по отношение на оценяването. Тя има максимуми в тези диапазони на оценката, където комисията оценява най-неточно и минимуми в диапазоните с най-точно оценяване. Обратната ѝ функция – *функцията на точността*, е характеристика на точното (увереното) оценяване – тя има максимуми в диапазона на оценката, където комисията оценява най-точно и минимум в диапазоните с най-неточно оценяване. Обхватът на баловите в скалата беше определен така, че да осигурява еднаква точност на оценяването независимо от това, в кой бал попада оценката. Това означава, че баловите около максимумите на функцията на точността, където оценяването е уверено, трябва да имат по-тесни граници, а тези около минимумите ѝ, където оценяването е с по-голяма грешка – по-широки граници.

За определяне на баловите граници беше използвана интегрираща програма, реализираща метода на трапеците²⁾, чрез която беше изчислен интегралът на функцията на точността в интервала на оценките. Броят на баловите беше избран така, че средната големина на баловите в скалата да е равна на максимума на грешката на комисията.

Скалата е с балове с еднаква точност, ако интегралът на функцията на точността има еднаква стойност за всеки от баловите. При това обхватът на баловите ще варира от бал към бал, тъй като функцията на точността не е константна. Стойността на интеграла на функцията на точността за всеки от баловите беше определена като отношение на интеграла на функцията за це-

лия интервал на оценките към броя на баловите в скалата. Долната граница на обхвата на максималния бал беше определена при известна горна граница 6,00 и вариране на долната граница до получаването на определената преди това стойност на интеграла за един бал. Долната граница на обхвата на максималния (последния) бал е горна граница на обхвата за предпоследния бал, чиято долна граница на обхвата беше определена по същия начин, и т.н. до достигане на долната граница на интервала на оценките 2,00.

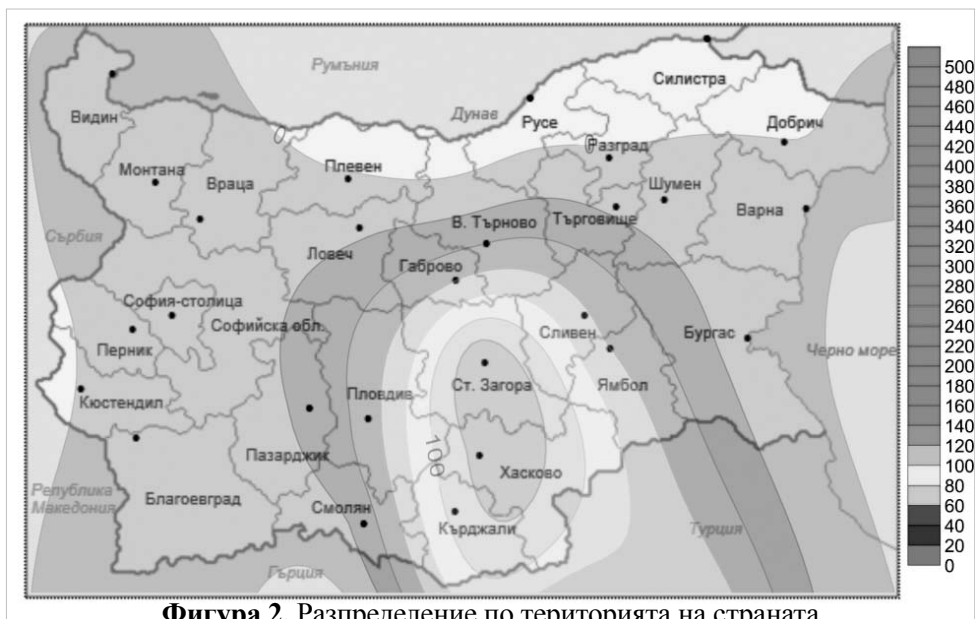
Резултати и обсъждане

Тракийски университет е базиран в три областни града в югоизточната част на страната – Стара Загора, Хасково и Ямбол. Разположението му определя и районите с преобладаващ интерес към специалностите в Университета, в частност към тези от тях, за които се кандидатства с изпит по Обща езикова култура (Фигура 1). От фигурата се вижда, че основният поток кандидат-студенти с изпит по Обща езикова култура в Университета е от региона Стара Загора, Хасково, Кърджали и Ямбол, като постепенно намалява с отдалечаването от тази група области. От някои райони на страната – Североизточна България и крайните западни райони, няма кандидат-студенти, кандидатствали с изпит по Обща езикова култура. Отношението на кандидат-студентите мъже:жени е 12:88.



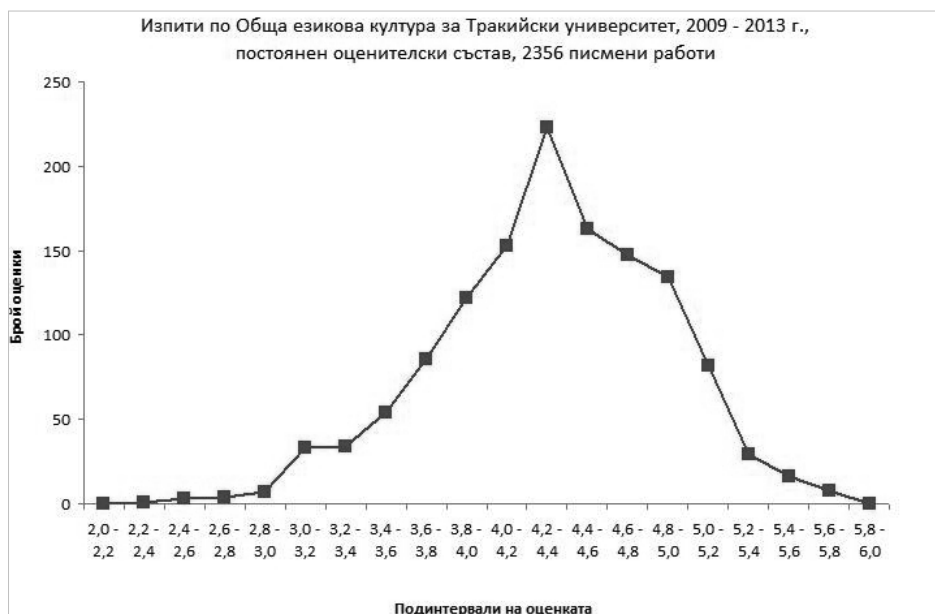
Фигура 1. Разпределение по територията на страната на броя кандидат-студенти, кандидатствали за последните 5 години в Тракийски университет с изпит по Обща езикова култура.

На фигура 2 е показано разпределението на средната оценка от изпитите по Обща езикова култура за Университета през изследвания петгодишен интервал. Вижда се, че най-подготвени са кандидат-студентите от област Варна, Ямбол, Габрово, Плевен. Средната оценка за мъжете е 3,77, за жените – 4,38.



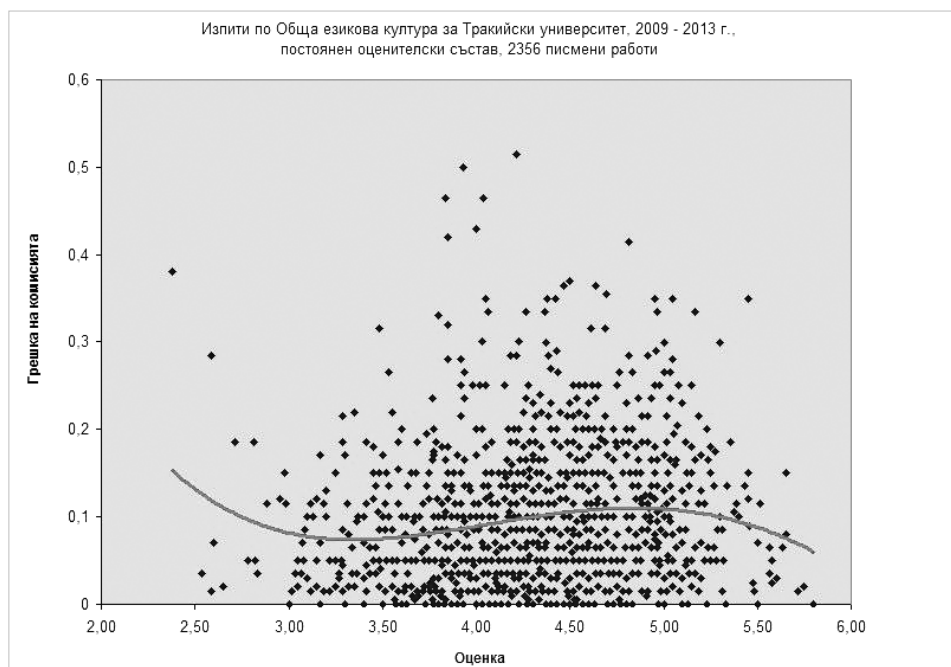
Фигура 2. Разпределение по територията на страната на средната оценка от изпитите по Обща езикова култура за Университета за последните 5 години

На фигура 3 е показано разпределението на оценките от 15-те изпита, обхванати в изследването. В интервала от оценки 3,40 – 5,20, представляващ 30% от интервала на оценките, попадат 90,2% от всички оценки. Поради грешката, неизбежно съпътстваща процеса на оценяване, високата концентрация на оценките води до висок риск от некоректно класиране на кандидат-студентите с близки оценки.



Фигура 3. Разпределение на оценките от 15-те изпита по Обща езикова култура, обхванати в изследването

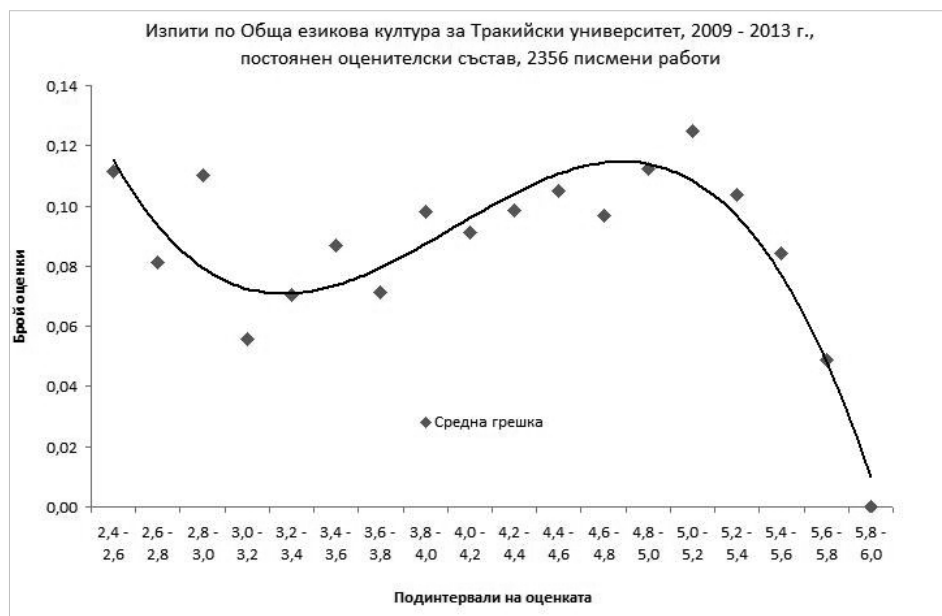
На Фигура 4 е показано разпределението в интервала на оценките на индивидуалните грешки (показани с точки) при оценяването на писмените работи от изпитите по Обща езикова култура. С линия е показана средната грешка на комисиите. Вижда се, че във важния за класирането интервал на оценката кривата има максимум (0,12) около петицата.



Фигура 4. Разпределение в интервала на оценките на грешката за петнадесетте оценителски комисии за изпитите по Обща езикова култура за интервала 2009–2013 г.

С точки са означени грешките на оценките на конкурсните работи. С плътната линия е показана средната стойност на грешката. Грешката е с максимум около петицата.

На фигура 5 е показана зависимостта на средната грешка на оценяващите комисии от оценката. Точките съответстват на средната грешка на комисията в подинтервали на интервала на оценките с дължина 0,2. Линията е регресионен модел на зависимостта средна грешка – оценка, наричана по-долу *функция на грешката*. Максимумът на кривата (0,12) е за оценки от подинтервала 4,80 – 5,00.



Фигура 5. Разпределение на средната грешка на комисиите в интервала на оценките (функция на грешката)

С точки са означени средните грешки в подинтервали от интервала на оценките с големина 0,2. С линия е означена регресионната крива на средната грешка на комисиите. Изключени са два от подинтервалите в началото на интервала на оценките – първият не съдържа оценки, а следващият е само с една оценка.

Негативното влияние на грешката на оценяване върху класирането може да бъде намалено, като на точковите оценки се съпоставят балове. Допълнително редуциране на негативното влияние на грешката би се постигнало, ако обхватите на баловите се направят пропорционални на големината на средната грешка на комисиите, т.е. на функцията на грешката в обхвата на бала. За изчисление на границите на обхвата на баловите беше използвана обратната функция на функцията на грешката, т.е. *функцията на точността*, изчислена като реципрочна стойност на функцията на грешката.



Фигура 6. Обратната функция на функцията на грешката
(функция на точността)

Показан е математическият израз на регресионния модел на функцията на точността (y) от оценката (x). Изключен е и последният подинтервал от интервала на оценката, тъй като не съдържа оценки.

Функцията на точността = $3,42708 \cdot \text{оценка}$

$-41,11829 \cdot \text{оценка}^2$

$+158,57131 \cdot \text{оценка}$

$+184,48476$.

Моделът е със сравнително висок коефициент на детерминация $R^2 = 0,66$, т.е. е адекватен на данните.

За определяне на границите на баловете, съобразени с точността на оценяване, беше използвано числено интегриране на функцията на точността, метод на трапеците²⁾. Интегралът на функцията на точността за целия интервал на оценките от 2 до 6 е 44,9994.

Броят на баловете беше подчинен, от една страна, на изискването за максимална детайлност на оценяването, т.е. максимален брой балове в скалата, а

от друга страна – на изискването средният обхват на бала да не е по-малък от максимума на средната грешка на комисиите. За определянето на броя балове в скалата общият обхват на интервала на оценките с големина 4 (горната граница 6,00 – долната граница 2,00) беше разделен на максимума на грешката на комисиите (0,12): $4/0,12 = 33,33$. Т.е. най-големият брой балове, отговарящ на посочените условия, е 33.

Съпоставянето на бал на точковата оценка е еднакво точен процес по цялата скала (рискът от грешка е еднакъв), ако интегралът на функцията на точността има една и съща стойност в обхвата на всеки от баловете. Стойностите на функцията на точността се изменят в диапазона на оценките, което при еднаква стойност на интеграла на функцията на точността за всеки от баловете води до различно широки обхвати на баловете от скалата. Стойността за бал на интеграла на функцията на точността се получава, като се раздели стойността на интеграла на функцията на точността за целия интервал на оценките на броя на баловете в скалата: $44,9994/33 = 1,3636$. За да бъде интегралът на функцията на точността с еднаква стойност 1,3636 за всеки от баловете, обхватите им трябва да са по-тесни при по-големи стойности на функцията на точността в интервала на бала (участъци с малка грешка на оценяване) и по-широки при малки стойности на функцията на точността (с голяма грешка на оценяване). Границите на обхватите на баловете бяха получени последователно, като се започна с максималния бал, чиято горна граница е 6,00. Чрез вариране на долната граница на интегриране на функцията на точността в обхвата на максималния бал беше подбрана такава долна граница (5,946, таблица 1), за която стойността на интеграла в обхвата на бала (1,4041) е максимално близка до 1,3636. Получената по този начин долна граница на обхвата на последния бал е горната граница на обхвата за следващия бал. Този процес беше приложен за изчисляването на границите на всички балове от скалата. Границите на обхвата са приведени в първите две колони на таблица 1. В третата колона е дадена ширината на обхвата на бала – разликата между границите му. Големината на бала е приведена в четвъртата колона. Максимално близката до 1,3636 стойност на интеграла на функцията на точността в рамките на бала, достигната в процеса на търсене на долната граница на обхвата (с точност до 3-тия знак след десетичната запетая) е дадена в петата колона. В последната колона е дадено разпределението на максималните оценки на 1299-те участници в конкурсните изпити за изследвания интервал време.

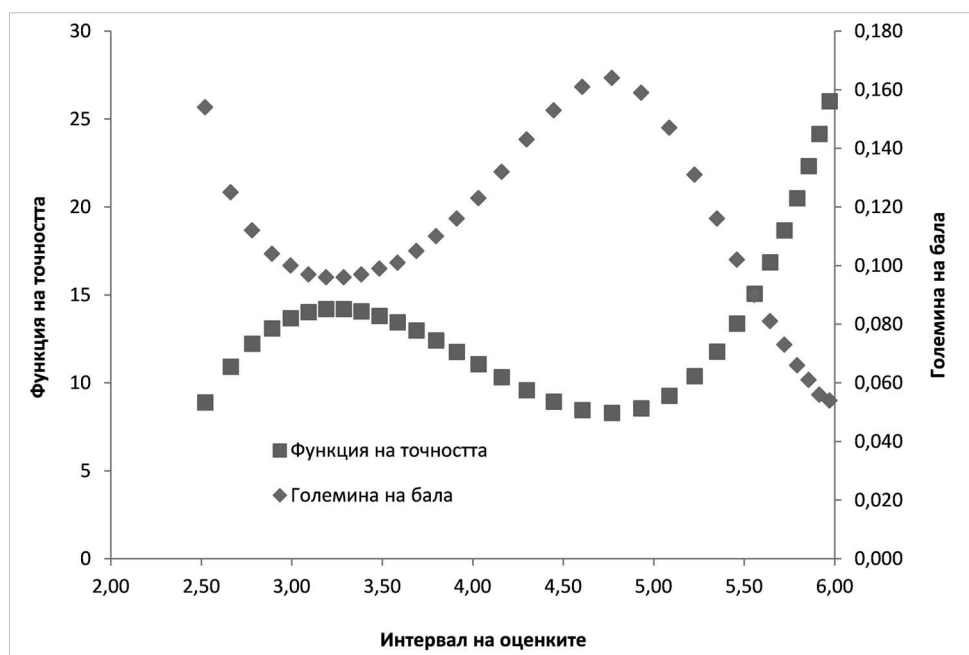
Средната ширина на баловете в скалата е 0,120. Балът с максимална ширина (0,164) на обхвата е „4,77“, с граници на обхвата от оценка 4,69 до 4,85. Всички точкови оценки, попадащи в този обхват, получават бал „4,77“. С минимална ширина (0,054) е последният бал „5,97“.

Таблица 1. Балове и балови граници на скала с балове с еднаква грешка на експертното оценяване, валидна за експертното оценяване на изпита по Обща езикова култура в Тракийски университет. Общата стойност на интеграла на функцията на точността за интервала на оценките от 2 до 6 е 44,9994. 1/33-та част от него е 1,3636.

Обхват на бала		Ширина на обхвата на бала	Бал (средна обхвата)	Най-близка стойност на интеграла до 1,36362	Брой на точкови оценки в обхвата на бала
Долна граница	Горна граница				
2	2,444	0,444	2,22	0,9524	1
2,445	2,599	0,154	2,52	1,3620	3
2,600	2,725	0,125	2,66	1,3616	3
2,726	2,838	0,112	2,78	1,3665	3
2,839	2,943	0,104	2,89	1,3603	2
2,944	3,044	0,100	2,99	1,3660	11
3,045	3,142	0,097	3,09	1,3592	15
3,143	3,239	0,096	3,19	1,3610	14
3,240	3,336	0,096	3,29	1,3617	20
3,337	3,434	0,097	3,39	1,3632	21
3,435	3,534	0,099	3,48	1,3660	27
3,535	3,636	0,101	3,59	1,3564	27
3,637	3,742	0,105	3,69	1,3608	42
3,743	3,853	0,110	3,80	1,3637	75
3,854	3,970	0,116	3,91	1,3635	71
3,971	4,094	0,123	4,03	1,3592	84
4,095	4,227	0,132	4,16	1,3612	111
4,228	4,371	0,143	4,30	1,3698	167
4,372	4,525	0,153	4,45	1,3656	128
4,526	4,687	0,161	4,61	1,3618	133
4,688	4,852	0,164	4,77	1,3618	113
4,853	5,012	0,159	4,93	1,3622	102
5,013	5,160	0,147	5,09	1,3643	66
5,161	5,292	0,131	5,23	1,3606	26
5,293	5,409	0,116	5,35	1,3667	11
5,410	5,512	0,102	5,46	1,3640	10
5,513	5,603	0,090	5,56	1,3568	7

5,604	5,685	0,081	5,64	1,3647	3
5,686	5,759	0,073	5,72	1,3624	2
5,760	5,826	0,066	5,79	1,3523	1
5,827	5,888	0,061	5,86	1,3616	0
5,889	5,945	0,056	5,92	1,3524	0
5,946	6,000	0,054	5,97	1,4041	0

Фигура 7 илюстрира обратната зависимост между ширината на обхвата и функцията на точността.



Фигура 7. Зависимост между функцията на точността и ширината на обхвата на бала – когато едната нараства, другата намалява и обратно

Калкулирането на точковите оценки в балове е лесно практически реализируемо.

Изводи

Работата на комисиите от експерти, оценяващи писмените изпити по Обща езикова култура, може да бъде оптимизирана, за да се намали грешката при оценяването и като резултат да се постигне максимално справедливо класиране.

За целта е необходимо:

- да се събират редовно данни за оценяването на членовете на комисията;
- да се доведе до знанието на всеки член от комисията резултатът от изследването за качеството на оценяването му;
- за арбитри да се привличат само членове на комисиите, които имат малка индивидуална грешка;
- да се намали грешката на комисиите чрез:
 - разработване на подробни критерии за оценка, разхвърлящи оценките максимално равномерно по скалата на оценките;
 - ограничена селекция на членския състав на комисията в зависимост от индивидуалната грешка на оценяване с цел минимизиране на грешката на комисията като цяло;
 - използване на скала с балове, при което се намалява рискът от несправедливо класиране на кандидат-студенти с близки оценки в резултат на грешката на оценяване;
 - оптимизиране на баловите в скалата така, че обхватът им да е пропорционален на грешката на комисията.

БЕЛЕЖКИ

1. Доцимология: <http://www.encyclo.co.uk/webster/D/104> Последно посетен: 13.6.2013 г.
2. Числено интегриране: http://people.hofstra.edu/stefan_waner/realworld/integral/integral.html Последно посетен: 13.6.2013 г.

STATISTICAL ANALYSIS OF THE GENERAL LANGUAGE CULTURE EXAMINATIONS FOR STUDENT'S ADMISSION IN TRAKIA UNIVERSITY, BULGARIA

Abstract. This work discusses the problems of docimological nature concerning the heuristic assessment of the works from written examinations in General language culture for the Thrakia university, the scores of which are used for the university applicants admission ranking. Due to the important life consequences of the assessment for the university admission purposes possible approaches to reduce the negative impact of the assessment error on the classification of students with similar grades are discussed. In this regard 33 grade scale is proposed whose grades have a range proportional to the distribution of the error in the range of estimates. Using such a scale whose grades depend on the error of estimation, makes the process of assessment equally accurate which reduces the risk of incorrect classification.

Assic. Prof. Nikolay Takuchev

✉ Trakia University
Stara Zagora, Bulgaria
E-mail: npt@uni-sz.bg