• Задачи •
• Problems •

ОСНОВНИ ПРОЦЕДУРИ ПО СТАНДАРТИЗАЦИЯТА НА ТЕСТОВЕ ОТ КАНДИДАТСТУДЕНТСКИТЕ ИЗПИТИ ПО ФИЗИКА ВЪВ ФИЗИЧЕСКИЯ ФАКУЛТЕТ НА СОФИЙСКИЯ УНИВЕРСИТЕТ[1)]

Димитър МЪРВАКОВ, Мая ГАЙДАРОВА, Клавдий ТЮТЮЛКОВ
Софийски университет "Св. Климент Охридски"

Резюме. Тази статия анализира тестовете по физика, които се използват като приемни изпити във Физическия факултет на Софийския университет "Св. Климент Охридски". Това изследване описва процедурите за конструиране на стандартизирани тестове, които могат да се използват като приемни кандидатстудентски изпити в Университета. На тази основа могат да се направят изводи за съвременното състояние на обучението по физика в българското училище. Резултатите от участието на 130 кандидатстуденти са представителни за цялата страна, защото тези кандидати са получили средното си образование в училища от различни райони на България.

Keywords: standardization procedure, admission tests, statistical methods, physics

Постановка на проблема

Приемните изпити за Физическия факултет на Софийския университет "Св. Климент Охридски" се състоят от тест със задачи с избираем отговор (т.нар. затворени тестове) и количествена задача, която проверява в по-голяма дълбочина равнището на физичните знания от задължителния училищен курс по физика. Задачите се дават съобразно предварително обявена програма в справочника за кандидат-студенти. В последните години форматът на изпита има някои промени — тестът се удължи от 15 на 20 задачи, като се запази броят на избираемите отговори. Удължаването на теста повишава неговата надеждност и намалява вероятността за случайно отгатване на няколко верни отговора. Това налага промяната на границата за успешност (среден три), която трябва да бъде по-висока от точките за най-голямата вероятност при отгатване на определен брой задачи. Тази вероятност пресмятаме с формулата за биномно разпределение

$$W_N(n) = \frac{N!}{n!(N-n)!} p^n q^{N-n} \quad ,$$

където N е дължината на теста, n е броят задачи, които случайно могат да се отгатнат, независимо от мястото им в теста, p е вероятността за случайно отгатване на верния отговор в една задача, а $q = 1 - p$. При дължина на теста 15 задачи с 5 възможни отговора Таблица 1 представя вероятностите за случайно отгатване на верните отговори на n задачи.

Таблица 1. Вероятности W_n за случайно отгатване на верни отговори при тест с 15 задачи с 5 възможни отговори

n	0	1	2	3	4	5	6	7	8	9
W_n	0,035	0,132	0,231	0,250	0,188	0,103	0,042	0,009	0,002	0,001

Най-голяма е вероятността за случайното отгатване на 3 задачи, всяка от които носи по 2 точки, или общо 6 точки. Затова границата за успешност е 12 точки. При дължина на теста 20 задачи с по пет отговора вероятностите се представят в Таблица 2.

Таблица 2. Вероятности W_n за случайно отгатване на верни отговори при тест с 20 задачи с 5 възможни отговори

n	0	1	2	3	4	5	6	7	8	9
W_n	0,012	0,058	0,137	0,205	0,218	0,176	0,109	0,055	0,022	0,007

Най-голяма е вероятността за случайното отгатване на четири произволни задачи от теста, което означава, че около 22% от извадката при голям брой кандидат-студенти при случайно посочване на верния отговор имат възможноста да отгатнат 4 задачи, което им носи 8 точки. Те също са под границата за успешност от 12 точки.

Цели на изследването

Основна цел на многогодишните анализи на резултатите от приемните изпити е организиране на база данни от задачи, които притежават необходимите качества за съставяне на стандартизирани тестове. Основание за прилагане на статистическите методики за анализ ни дава относително голямата извадка от кандидат-студенти, от чиито работи емпирично се обособява база данни за обработване. За съжаление тази извадка намалява всяка година, като от над 200 кандидат-студенти през 2006 г. стига до 68 през 2009, което понишава презентативността на изследванията. Извадката е представителна за страната, защото включва представители от всички области.

Част от кандидат-студентите са се явявали и на държавен зрелостен изпит по физика (ДЗИ) по физика, което позволява да се направят някои изводи за валидността на теста по отношение на други тестове. За тази цел се правят и допълнителни измервания на тестовете в последните класове на гимназиите, основно във физическата паралелка на Националната природоматматическа гимназия.

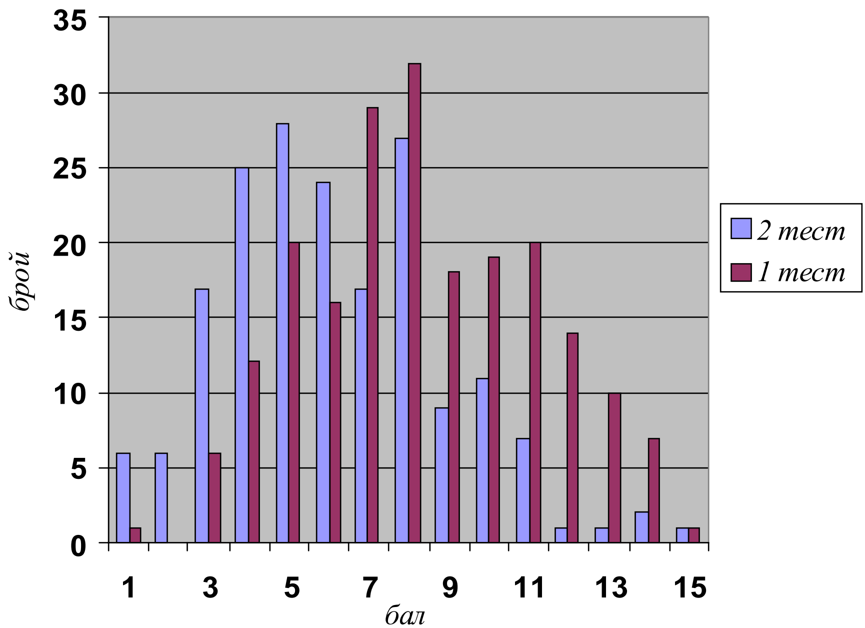
Измерванията на постиженията на тестираните кандидат-студенти и ученици по отделните задачи позволява също така да се направят изводи за трудностите по усвояването на задължителния учебен материал по различните учебници и да се анализират основни грешки в познанията и уменията на учениците от различни възрастови групи.

Методика на изследването

Анализ на теста като цяло

Анализът на теста започва с проверка на разпределението. Винаги тестовият бал има нормално разпределение (Фиг. 1), което се проверява теоретично с критерия на Пирсън [1-4]. На Фиг.1 е показано разпределението за 2006 г. Изчисляват се коефициентите на асиметрия и ексцес, които представят степента на отклонение от симетричното Гаусово разпределение и неговата полегатост. Например за първото разпределение те са съответно

$$\alpha_1 = 0,32,\ \beta_1 = 3,28 \text{ и } \alpha_2 = 1,36,\ \beta_2 = 3,68 .$$



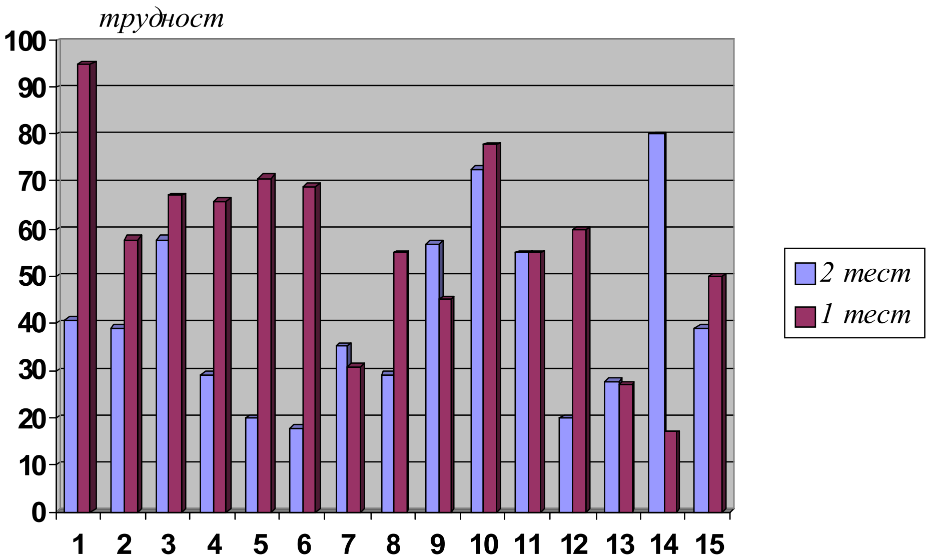
Фиг. 1. Брой кандидат-студенти/бал

Разпределението от първия изпит е с малка дясна асиметрия и близка до нормалната полегатост, а от втория — с по-голяма дясна асиметрия с бимодално разпределение и островърха полегатост. Средните стойности от тестовия бал за първия и втория изпит са съответно $\overline{x}_1 = 16,55$ и $\overline{x}_2 = 12,34$, което показва, че неявилите се на втория изпит кандидат-студенти имат високи резултати от първия изпит и те са променили качествено състава на първоначалната извадка. Нормите, изведени за първия тест, ще имат по-голяма статистическа достоверност, защото извадката по-добре отразява характеристиките на генералната съвкупност.

Оценяваме надеждността на теста (по метода на разделяне на теста на две половини), т.е. точността, с която оценяваме тестираните и стандартната грешка на измерването. Например при тест 1 при стандартно отклонение $\sigma = 5,93$ и коефициент на надеждност $R = 0,92$ стандартната грешка е $s = \sigma\sqrt{1-R} = 1,68$, т.е. точността на оценяване на знанията на всеки студент са в интервала ± 2 точки около неговия бал. Като се има предвид, че на 0,25 от бала отговарят 3 точки от теста можем да твърдим, че изпитваните са оценявани с грешка, по-малка от 0,2 от оценката им по шестобалната система. За втория тест при $\sigma = 2,76$ и по-малък коефициент на надеждност $R = 0,78$ за стандартна грешка се получава $s = \sigma\sqrt{1-0,78} = 1,3$, което дава по-добра точност от първия тест. Това се дължи вероятно на по-хомогенната извадка при втория тест.

Анализ на задачите

Анализът на трудността на задачите за двата теста (Фиг. 2) показва, че за първия тест всички задачи, с изключение на тези с номера 1 и 14 (първата е много лесна, а втората — много трудна), са в границите на препоръчителната трудност (20% — 80%). За втория тест над препоръчителната трудност са 5-та, 6-та и 12-та задачи (но при по малка извадка). Такива задачи се преработват и отново се апробират или се отстраняват от базата данни.



Фиг. 2. Трудност на задачите

Всяка решена задача от теста се оценява с еднакъв брой точки, което предполага, че те трябва да са приблизително с еднаква трудност. Това се проверява за съжаление емпирично след провеждане на изпитите поради липса на стандартизирани тестове. И в двата входящи теста има задачи, проверяващи усвояването на едни и същи знания и умения по физика. Една от целите на настоящото изследване е да установи различията в трудностите на такива задачи и да анализира причините за това различие. За осъществяването на това се анализират отговорите на кандидат-студентите, взели участие и в двата изпита и решавали всяка от двайсетте задачи от теста.

За целта се използва статистическият критерий χ^2 на Пирсън за проверка на разпределението на данни от четириполева таблица (Таблица 3), която съдържа следните емпирични честоти: *a* — получената честота на лицата, решили втората, но нерешили първата задача; *b* — получената честота на лицата, решили и двете задачи; *c* — получената честота на лицата, нерешили и двете задачи; *d* — получената честота на лицата, решили първата, но не и втората задача.

Таблица 3

	Нерешил А	Решил А
Решил В	a	b
Нерешил В	c	d

От постановката на проблема е ясно, че са съществени честотите *a* и *d*, които отчитат различията при решаването на задачата, проверяваща едни и същи знания в двата варианта А и В на приемните изпити.

Очакваните честоти за нормалното разпределение се получават като средноаритметично от честотите a и d,

$$\text{т.е.}\quad f_1 = f_2 = \frac{a+d}{2} \quad .$$

Следователно за χ^2 получаваме

$$\chi^2 = \frac{\sum_{i=1}^{2}(f_i - e)^2}{f_i} = \frac{\left(\frac{a+d}{2} - a\right)^2}{\frac{a+d}{2}} + \frac{\left(\frac{a+d}{2} - d\right)^2}{\frac{a+d}{2}} = \frac{(a-d)^2}{(a+d)} \quad .$$

За всяка съответна двойка задачи от двата теста сравняваме получения χ^2 с критичния коефициент, чиято стойност за избраната статистическа вероятност $\alpha = 0,05$ и степен на свобода $k = 1$ е $\chi^2_{0,05,1} = 3,84$. Ако $\chi^2 < \chi^2_{0,05,1}$ разпределението е нормално и двете задачи са с еднаква трудност. Ако $\chi^2 \geq \chi^2_{0,05,1}$, задачите са с различна трудност. Степента на отклонение от критичната стойност показва и разликите в трудността на задачите.

Определихме 12 двойки задачи от двата приемни изпита за 2008 г., които по експертна преценка проверяват едно и също знание или умение, необходимо за решаването на задачата. Това са двойките задачи със съответните номера в теста, показани в Таблица 4.

Таблица 4

8. юли	1	3	4	5	7	10	12	13	14	15	18	19
18. юли	1	3	4	6	5	10	11	13	14	15	18	19
	3,24	28,44	1,50	16,33	13,51	18,18	5,56	7,53	0,07	0,81	20,16	1,8

Пресметнатите стойности на критерия на Пирсън χ^2 за всяка от двойките задачи по описаната по-горе методика са дадени в таблицата. По-малка стойност от критичната $\chi^2_{0,05,1} = 3,84$ имат задачите 1-1, 4-4, 14-14, 15-15, 19-19. Те са с еднаква трудност за извадката, която е решавала и двата варианта на теста.

Задачите, които теоретично се очаква да имат приблизително еднакви трудности, емпирично се оказват с много различни трудности. Това не зависи от извадката, тъй като тестовете се провеждат с едни и същи лица. Отклоненията се дължат на различни причини, част от които бяха изброени по-горе. От гледна точка на надеждността на теста като измерителен инструмент, която се влияе от скалирането, следва да се присъждат и различен брой точки на изследваните двойки задачи, които показват значителни разлики с критичния коефициент на Пирсън χ^2. Това показва, че процедурите по стандартизация на тестовете по физика за приемните изпити изискват достатъчно дълго време, като се има предвид, че апробацията им засега се осъществява само по време на самите изпити.

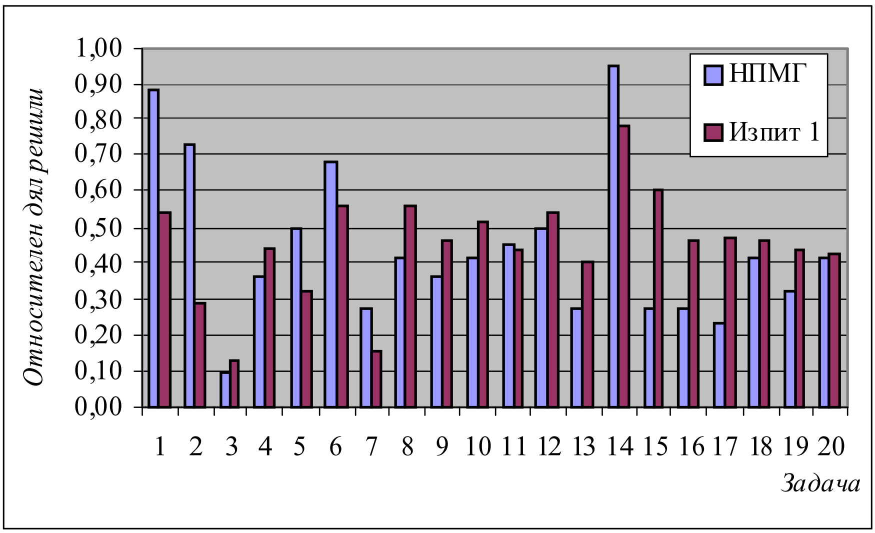
Като метод за проверка на валидността на тестовете се провежда и паралелната им апробация във физическата паралелка на Националната природоматематическа гимназия с ученици от 12. клас, които са изучили учебното съдържание от обявената програма в кандидатстудентския справочник. В това изследване се сравнява и трудността на задачите, емпирично установена за двете извадки — кандидат-студенти и ученици от Националната природоматематическа гимназия.

За сравняването на трудността на задачите в двете различни извадки (кандидат-студенти и ученици от Националната природоматематическа гимназия) използваме метода за сравняване на относителни части. За всяка задача се определят относителните части p_1 и p_2 на решилите я от двете извадки. Разликата от относителните части $d = p_1 - p_2$ показва различията в трудността на задачата за всяка от групите. Чрез статистическия критерий на Стюдънт

$$t_{st} = d/s_d \text{ , където } s_d = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \text{ , а } n_1 = 22 \text{ и } n_2 = 68$$

са броят тестирани съответно от Националната природоматематическа гимназия и първия приемен изпит (8 юли), се проверява дали разликите в трудността за двете групи са статистически значими при избрана вероятност $\alpha = 0.05$. Получената величина сравняваме с критичната стойност на t-критерия на Стюдънт за съответната степен на свобода $k = n_1 + n_2 - 2 = 22 + 68 - 2 = 88$ от таблицата $[1]$. Определяме $t_{88,0.05} = 1{,}98$. Ако $t_{st} < t_{88,0.005}$ няма статистически значима разлика между трудността на задачата за двете извадки.

На хистограмата (Фиг. 3) са показани относителните дялове за задачите. В Таблица 5 са пресметнати коефициентите на Стюдънт за разликите в тези дялове. От таблицата се вижда, че само при пет от задачите от първия вариант — 1, 2, 14, 15 и 17 има статистически значима разлика в трудността за двете групи. Това са задачи от раздел механика и раздел вълни и частици. При първите различията в трудността се дължат вероятно на изучаването на механиката в 10-ти клас, т.е. част от материала е забравен, но за различията в трудността на задачите от втория раздел няма задоволително обяснение.



Фиг. 3. Относителен дял на решилите задачи 1-20

Една част от изследванията са свързани със сравняване на резултатите от получените оценки от матурата по физика и кандидатстудентския изпит. За тази цел пресмятаме коефициента на Пирсън за метрично скалирани величини $R=\frac{\sum_{i=1}^{25}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sigma_x\sigma_y\left(n-1\right)}$, където x_i и y_i са съответно оценките на $n=25$ кандидатстуденти, явили се на приемен изпит и матура, $\bar{x}$ и $\bar{y}$ са средните стойности, а σ_x и σ_y са средноквадратичните им стойности.

Таблица 5. Коефициенти на Стюдънт

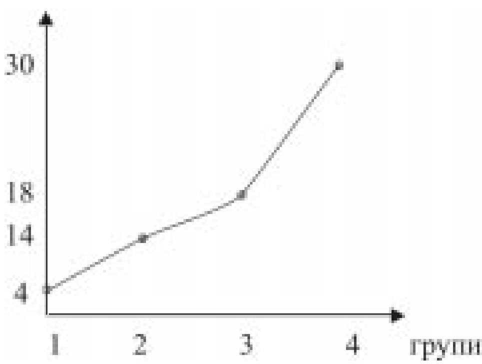
N	1	2	3	4	5	6	7	8	9	10
t_{st}	3,69	4,01	0,54	0,67	1,49	1,03	1,15	1,24	0,84	0,82
N	11	12	13	14	15	16	17	18	19	20
t_{st}	0,16	0,32	1,16	2,48	2,95	1,69	2,21	0,41	0,95	0,08

За коефициента на корелация на оценките от матурата и приемните изпити получихме $R=0,68$, което показва значителна корелация между тях. Въпреки известната разлика в обема и естеството на проверяваното съдържание при приемния изпит и матурата, която не включва разделите механика и топлинни явления, кандидат-студентите, които са се явили и на матура показват доста сходни постижения на двата изпита. Това според нас означава, че подготовката само на някои раздели от физиката (електричество и магнетизъм, трептения и вълни, оптика и атомна и ядрена физика) имат значение и за доброто представяне по механика и топлинни явления.

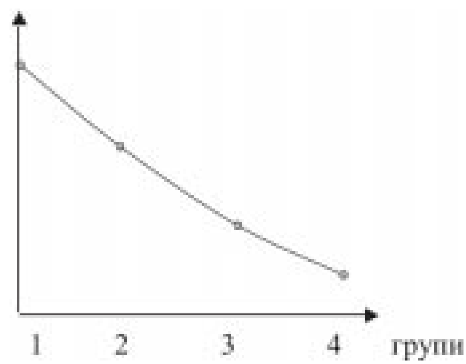
За изследване на ефективността на дистракторите отново използваме статистическия критерий χ^2 на Пирсън за проверка на разпределението. Наблюдаваните честоти за отговорите на всяка от задачите e_i (задачите са с пет отговора), се сравняват с теоретичните t_i, които при случайно отговаряне на въпроса са равни на средноаритметичното от сумата на всички отговори. По формулата $\chi^2=\sum_{i=1}^{5}\frac{\left(e_i-t_i\right)^2}{t_i}$ за всеки въпрос определяме χ^2 и сравняваме стойността с критичната стойност при избрана вероятност $\alpha=0.05$ и степен на свобода 4 (броят на отговорите минус 1). От таблицата $[1]$ определяме $\chi^2_{4,0.05} = 9,49$. Ако $\chi^2_{4,0.05} < \chi^2$, на въпроса не е отговаряно случайно и той е добре разбран.

При изследването на дистракторите установихме, че при всички задачи те “работят”, т.е. посочвани са, като при задачи , напр. при 11 и 13 от първия вариант дистракторите съответно “д” и “в” са маркирани по един път. Като се вземе предвид неголямата извадка (68 души), това може да се приеме. За всички задачи стойността на коефициента на Пирсън χ^2 за разпределението на отговорите е по-голяма от $\chi^2_{4,0.05} = 9,49$, което означава, че отговорите не са посочвани случайно. Но за някои от задачите, напр. 3-та и 7-ма от първия изпит верният отговор — съответно “б” (16) и “а” (10) е посочван по-малко от друг дистрактор — съответно “г” (29) и “г” (28). Това може да се обясни с недоразбиране на условието на задачата или дистракторите. И двете задачи е необходимо да се доуточнят и променят, за да влязат като задачи с добри характеристики в базата данни.

За изследването на качествата на предварително структурираните отговори на задачите поотделно използвахме техните характеристични криви (Фиг. 4). Те се изготвят, като извадката се раздели на четири еднакви групи — силна, висока средна, ниска средна и слаба и се определи броят на учениците, решили задачата от всяка група. Кривата, която свързва четирите точки, трябва да е монотонно растяща, за да има задачата достатъчна дискриминативна сила, т.е. да отделя силните от слабите кандидат-студенти. Това е важно за нормативните тестове, какъвто е кандидатстудентският. Обратно, характеристичната крива на дистрактора (на Фиг. 5 е представена само една от четирите за всяка задача) трябва да бъде намаляваща. За почти всички задачи това е изпълнено. При останалите се налага преработване или смяна на дистракторите.



Фиг. 4



Фиг. 5

Процедурата по стандартизация на тестовите задачи се нуждае от повече време за апробации и изследване на качествата на теста и задачите, но е необходима за осигуряване на по-надежден и качествен инструментариум за обективно измерване на знанията и уменията на кандидат-студентите. Усъвършенстването на тестовете и подобряването на точността им повишават и доверието в обективността на изпитите.

Това изследване се проведе с помощта на фонд "Научни изследвания" към Софийския университет, за което изказваме благодарност.

БЕЛЕЖКИ

1. Доклад на 43-та Национална конференция на учителите по химия в Ловеч, 26-28 ноември 2009 г.
2. Тестове за приемни изпит са публикувани в кн. 5 (2008) на сп. *Физика*.

ЛИТЕРАТУРА

1. **Клаус, Г., Х. Ебнер**. Основи на статистиката за психолози, педагози и социолози. Наука и изкуство, София, 1971.
2. **Гласс, Д., Д. Стенли**. Статистические методы в педагогике и психологии. Прогресс, Москва, 1976.
3. **Bachman, L.F**. Fundamental Considerations in Language Testing. Oxford University Press, Oxford, 1990.
4. **Ravid, R**. Practical Statistics for Educators. University Press of America, Lanham, 2005.

STANDARDIZATION PROCEDURE OF SOFIA UNIVERSITY ADMISSION TESTS IN PHYSICS

Abstract. The aim of the present survey is to describe the content of the Sofia University admission test in Physics, to analyze the quality of the problems and a posteriori the character of the test, as well as to make some evaluation and conclusions related to the education of physics in Bulgarian schools. This survey is

a part of the standardization procedure of admission tests to secondary schools and universities in accordance with the contemporary requirements in the educational system. The ground on which the statistical methods of analysis are applied is based on an extract of more than 130 candidate students whose tests empirically originated a data base. The extract is representative for the country because it includes representatives from all areas.

✉ **Dr. Dimitar Marvakov**,
Department of Theoretical Physics,
University of Sofia,
5, James Bourchier Blvd., 1164 Sofiam BULGARIA
E-Mail: marvakov@phys.uni-sofia.bg

✉**Dr. Maya Gaydarova,**
Dr. Klavduy Tutulkov,
Department of Physics Education,
University of Sofia,
5, James Bourchier Blvd., 1164 Sofia, BULGARIA
E-Mail: mayag@abv.bg
E-Mail: klavdiyt@abv.bg