MULTIPLE-CHOICE QUESTION EXAMINATIONS IN PRE-CLINICAL COURSES AS POTENTIAL ENDPOINTS FOR ANALYSIS OF QUALITY OF MEDICAL SCHOOL ADMISSION CRITERIA

M. Panczyk, J. Belowska, J. Gotlib

Division of Teaching and Outcomes of Education, Faculty of Health Sciences, Medical University of Warsaw (POLAND)

Abstract

Introduction

Various forms of admission tests are used in most western countries to select the best candidates for medical schools. Predictive ability of selected examination criteria with reference to student's results obtained in pre-clinical and clinical courses constitutes an important factor that determines good selection of candidates.

Aim of study

Assessment of the quality of final tests in physiology and pathophysiology courses as well as retrospective analysis of the medical school admission criteria with reference to the results of pre-clinical education.

Materials and Methods

862 full-time and part-time Warsaw Medical University students, including 416 students admitted in the academic year 2008/09 and 446 students admitted in the academic year 2009/10 to the two medical faculties. The results of the biology, chemistry, and physics exams were analysed to assess the admission criteria. The results of two physiology exams and four pathophysiology exams were used to analyse the quality of multiple-choice question (MCQ) examinations in pre-clinical courses.

The quality of the exams was assessed by establishing the level of simplicity and differentiating power of particular questions and the reliability of the exam was assessed by estimating Cronbach's α coefficient. Moreover, the analysis involved the correlation between the results of physiology and pathophysiology exams as well as between particular results of these exams and scores obtained by a candidate during the admission procedure. The Pearson product-moment correlation coefficient was used to assess the correlation.

The significance level for all analyses was established at $p<0.05$.

Results

A detailed analysis of the quality of physiology and pathophysiology exams demonstrated high reliability of all MCQ tests ($\alpha > 0.08$). Furthermore, neither the number of distractors, nor the number of test versions significantly influenced the quality of these tests. Optimisation of examination tools consisting in elimination of questions with a negative correlation improved the quality of tests (on average by +0.02). Difficulty of particular tests was diverse and fell within 0.55 and 0.80. A correlation analysis showed a statistically significant relation between the physiology exam results and pathophysiology exam results (r value between 0.521 and 0.624, $p < 0.001$). The total reliability of pre-clinical exams was also high and fell within $\alpha = 0.872 – 0.922$.

Assessment of the selective ability of the adopted admission criteria demonstrated that, while in the case of students admitted to the full-time programme the total score obtained in the three key courses (biology, chemistry and physics) had a positive correlation with the results of pre-clinical exams ($p<0.01$), there was no such correlation in the case of part-time students. Moreover, the analysis of individual criteria showed that chemistry was the best independent selection criterion, whereas biology and physics indicated low or no correlation.

Conclusions

The accuracy of admission of appropriate candidates for medical school should be evaluated on several levels during the course of studies as well as when monitoring a medical career of a graduate.

ISBN: 978-84-617-0557-3

There is a need to strive for a proper admission procedure that would allow for selection of the best candidates for prospective doctors who would efficiently learn complex pre-clinical and clinical skills.

Keywords: medical education, admission criteria, pre-clinical education, educational measurement, multiple-choice questions (MCQs).

1 INTRODUCTION

Medical studies have always been immensely popular with the high school graduates. There are on average, 3-4 people per one place there. Universities accept candidates for the first year studies on the basis of a competition, applying various strategies so as to select the best ones. Since 2005, the mechanism of selection in Poland has been based on the results of national matura exam (an equivalent of A-levels in UK, Baccalauréat in France or Abitur in Germany) [1]. The result of matura obtained by a graduate becomes the basis on which he may be accepted at one of the 11 medical universities, including the Medical University of Warsaw (WUM). Additionally, every university has the right to establish their own criteria that would be included during the recruitment process, by deciding which high school subjects and at what level (basic or extended) will become the basis of selection. This means that individual medical universities may apply different recruitment criteria. Better universities, which a much higher number of candidates apply to, have the possibility to use stronger discrimination criteria and the selection of the best candidates from among all who decided to apply, may be more effective. In many countries there are similar systems of recruitment, based on evaluation the candidate by grade point averages (GPAs) in the USA or General Certificate of Education Advanced Level (A-level) in the UK, and aptitude tests. GPAs assess knowledge of academic subjects, such as chemistry and mathematics. Aptitude tests assess a mixture of reasoning, intelligence and culturally acquired knowledge, such as vocabulary and geography [2]. Within the recruitment process at WUM between the years 2008-2010, the results of national matura exam in biology, chemistry and physics were taken as the basic criterion when accepting an application for medical studies [3]. Since medical studies are extremely demanding and the career path requires several years of learning prior to specialisation period and also constant further increase of qualifications, it is of essence that only the best and most motivated candidates be accepted. Due to the specificity of education in medicine and high costs, candidates ought to met very high demands, so that it could be stated with high probability that they will achieve the professional and educational success [4, 5].

In Poland, medical studies require 12 semesters divided into 6 years of studies. The first years are devoted primarily to learning the basic disciplines that include subjects of morphological and bio-chemical and physiological profile, such as histology and embryology, proper anatomy, biophysics and physiology. The third year of studies includes teaching pre-clinical disciplines with a group of subjects from pathology: pathological anatomy or pathophysiology. Physiology or pathophysiology realise the outcomes of B block: scientific basis of medicine to which 330 didactic hours are devoted (overall 25 ECTS points) [6]. Thus, learning in the first three years precedes the key elements for the future doctor, i.e. classes in clinical subjects. Knowledge and skills gained previously are to help solve practical medical problems and it is complemented with proper knowledge, specific to a given clinical discipline. It may thus be assumed that competences gained during the first years of studies are pre-requisite and a condition necessary for effective education in the clinical sciences in the future [7].

There were numerous works published within the last fifty years that concerned evaluation of recruitment systems of medical universities, based on different criteria of candidates' selection [4, 8-26]. One of the important aspects of this research is evaluation of predictive ability of the selected criteria of recruitment and its influence on the outcomes gained throughout studies in pre-clinical and clinical classes, also in relation to the points obtained in professional licence exams (e.g. USMLE, United States Medical Licensing Examination). According to the recommendations of Admissions to Higher Education Steering Group, an ideal and fair system of selecting candidates should, on the one hand, be based on the criterion of a competence assessment, referred to as cognitive achievements, and on the other hand, also non-cognitive features should be assessed [27, 28]. As Ferguson *et al.* emphasise, (2002) [9] it is necessary to carry out well-planned, thorough and long-term prospectuve research that would establish the real predictive value of the selected cognitive and non-cognitive elements in relation to forecasting the future fate of prospective physicians.

The aim of this study was to evaluate the quality of final exams testing physiology and pathophysiology, and retrospective analysis of selection criteria among the candidates for medical studies in relation to the outcomes of pre-clinical teaching.

2 MATERIALS AND METHODS

Data concerning 862 students of WUM were used in the retrospective analysis, including 416 students from the year 2008/09 and 446 students from the year 2009/10 who were both full-time and part-time students at two Faculties of Medicine. When analysing the criteria of candidates' selection, results of three entrance examinations were considered: biology, chemistry and physics. Whereas in the quality analysis in form of a multiple-choice question (MCQ) exam from the section of pre-clinical education, data from 2 exams in physiology and 4 exams in pathophysiology were used.

Raw outcome data were pre-developed using software TESTY version 7 (*Testy komputerowe*, Copyright © 1994-2014 by Sławomir Zalewski). Information uploaded to database was further processed using EXCEL (*Microsoft*, 2010) and exported to a statistical package of STATISTICA (*StatSoft Inc*., version 10, 2011), where it was processed even further. All software programmes were used according to WUM licence.

Evaluation of normal spread of parametres of individual exam results was assessed using a Shapiro-Wilk test and the analysis of resistance of data to the oultiers was performed using a Grubbs test. The quality of individual questions from testing exams was assessed by establishing a coefficient of easiness and discriminating replacement power, and in order to determine the level of reliability of the test, an Cronbach's α coefficient value was determined (Kuder-Richardson coefficient for the test comprising of a two-categorical position) [29]. In order to compare the equivalence of test variants in pathophysiology for the first and the second Faculty of Medicine, a non-parametric U Mann-Whitney test of medians comparison was applied. Additionally, there was also performed an analysis of correlation of outcomes in physiology and pathophysiology exams, and individual results of these exams with scoring obtained by a candidate during a recruitment process to university. An r-Pearson coefficient of linear correlation was applied.

The level of statistical relevance for all analyses was established at $p < 0.05$.

3 RESULTS

While analysing the results of individual editions of exams in physiology and patophysiology considering the character of variable distribution, an asymmetric (oblique) and, in most cases, deviating from the normal distribution of data set (kurtosis ≠ 0 and Shapiro-Wilk test, $p < 0.05$). Such type of variable distribution is typical for exams of selective character. Individual editions of exam tests differ in terms of a range of outcomes changeability, which is reflected in various values of variation coefficients, hiatus in the outcomes and the range of points obtained by students in individual years. While analysing individual parametres we may point out to the differentiated level of difficulty in consecutive years. All results are presented in Table 1.

The measure of internal consistency of the test questions is the reliability of the test. The most frequently used technique of evaluating the internal consistency for at least two factors is an Cronbach's α reliability coefficient. The analysis of individual editions of the exam points out to high reliability of a test question set in consecutive years – α coefficient above 0.8. Detailed evaluation of questions allowed to extract test tasks with a negative correlation and / or insufficient value of discriminating power. Following the elimination of these questions from the exam set, the overall reliability of tests was increased and after optimization, α coefficient increased by 0.02 on average. The coefficient of test difficulty in consecutive years oscillated around 0.55 – 0.80 (Table 1).

Since there are two individual exams carried out in pathology, depending on the faculty the students are on, there was an analysis of equivalence of variants performed for these exams. Because the distribution of results in the compared groups differed to the normal one (Shapiro-Wilk test, $p < 0.05$), therefore a comparison test of medians was also applied. In 2010, the students of the sophomore year of Medical Faculty obtained significantly higher results in the exam in pathology as opposed to the results of students from the first year (medians 69.0 and 55.0 respectively, U Mann-Whiney test, $P< 0.001$). In case of exams carried out in 2012, there was also a statistically important difference observed, however, this time it was the students of the first year who gained higher results (U Manna-Whiney test, $P< 0.001$).

In order to determine the relationship between the results obtained by the students in physiology and pathophysiology, the analysis of correlations was performed. As presented in Table 2, statistically significant correlations with r-Pearson coefficient values of above 0.5 were obtained. Moreover, total reliability for exams in physiology and pathophysiology was calculated in consecutive years and the

value of α coefficient obtained was above 0.85. A sample graph of distribution for the analysed correlations was presented in Fig. 1.

In order to evaluate predictive ability of selective criteria for the candidates for medical studies, recruitment data was used. The degree of correlation of ranking points and score from the subjects criteria in biology, chemistry and physics was determined, with the points obtained by a student in pre-clinical education in physiology and pathophysiology. Assessment of selective skills of the accepted recruitment criteria shows that, if in case of full-time students the overall score of points in three subjects (biology, chemistry and physics) has a positive correlation with the results of preclinical exam tests ($P< 0.01$), a similar dependence was not observed in case of part-time students. Moreover, the analysis of individual criteria shows that, as an individual factor of selection, chemistry functions best, whereas biology and physics present only a very weak correlation or none at all. Details of the correlation analysis were presented in Table 3.

4 DISCUSSION

In order to select the best candidates for medical studies, in most Western countries different forms of admission tests are applied. Internationally, standardised admission tests, such as the Medical College Admissions Test (MCAT) [30], the Graduate Australian Medical School Admissions Test (GAMSAT) [8], the UK Clinical Aptitude Test (UKCAT) [31], are widely used. However, their added value to the selection process is yet to be clearly established [32, 33].

Until the year 2004 there was a uniform mechanism in Poland that would help select candidates, based on the national aptitude test, which was similar in structure to other exam tests of that kind, e.g. the American MCAT [30]. This exam would include the candidate's evaluation in terms of his / her knowledge and skills on the high school level in three subjects: biology, chemistry and physics. As of 2005, there has been a new mechanism of selection, based on the results of the national matura exam. The same range of subjects is assessed as in the previous system. There are over 2 thousand candidates applying for a place at WUM annually, 500 of whom has a possibility to study full- or part-time within the 6-year system of university education. Considering the fact that the number of candidates is four times the number of available places, it is necessary to implement a well-planned selection strategy and the university recruitment policy ought to be constantly controlled and evaluated.

In available literature we can find various methodological models used in evaluating criteria depending on their selective ability. In predictive studies it is determined how the applied criteria of selecting candidates relate to the future achievements of students of medicine. The reference point can be results of subject tests or average of overall score from the period of studies, or chosen thematic parts: basic sciences, preclinical and clinical sciences [8]. Another end point of analysis could be the result of the summarising exam or a licence exam of American type, such as USMLE [32], Canadian LMCC (Licentiate of the Medical Council of Canada) [34] or British MRCP (Membership of the Royal Colleges of Physicians) [35]. Finally, selective predictor in entrance examination could be evaluated with reference to specialisation exams, e.g. USMLE step 2 cs and step 3 [36].

To be able to use the results of subject tests in evaluating the predictive ability of recruitment criteria, they need to fulfil the conditions of high reliability and accuracy in their measurement of a student's achievements.

While analysing the parameters of descriptive statistics for individual editions of exams in physiology and pathophysiology, we could point out to the differentiated level of difficulty in consecutive years. Moreover, exams in pathophysiology organised twice a year for two different faculties are not equivalent. The above observations are evidence that the conditions of measuring in consecutive years of subject exams are not standardised. Therefore, the results of analysis of correlation with the criteria of recruitment for candidates may be subject to some error in the lack of uniformity in evaluation throughout the course of study.

Apart from equivalence in the test variation, its reliability is an important element, i.e. the measure of the internal consistency of a set of test questions, most frequently measured with the use of Cronbach's α coefficient. Quality of individual questions influences reliability of a test as well as balance in the selection of questions in terms of their difficulty and the range of scale in which the obtained results are measured. It is assumed that a 100-question test is reliable enough if α is not lower than 0.7. For the analysed test exams, high values of $\alpha > 0.80$ were obtained. With such > 0.80 values, random errors constitute less than 20% of variability of the obtained results and the

measurement in such conditions may be applied in individual differentiation [37]. However, it needs to be pointed out that high values of α coefficient do not solve the problem of exam reliability since high α value only means minimizing the influence of random errors on the obtained results, which does not give any certainty as to the existence of systematic errors that are sometimes serious [37].

When creating good tools for assessment it is important to determine the accuracy of measurement that answers the question: "what is being measured"? Accuracy here needs to be understood as usefulness of a given tool in evaluating a certain set of features and properties of the examinee [38]. Due to the fact that the effects of education achieved by students as part of their course in physiology are necessary to effectively evaluate issues in pathophysiology [6], a good measure of accuracy of should be a positive correlation between the results in both subjects. As was shown in Table 2, for all editions of exams in both subjects, there is a certain relevant statistical correlation and r-Pearson correlation coefficients are higher than 0.5. The obtained linear model of dependencies is of average quality because determination coefficient r^2 has values within 0.25. This means that around 25% of variability in evaluation of exam in pathophysiology may be explained by the results that a student achieved in physiology. Concluding, it could be stated that the result achieved by a student in pathophysiology exam is greatly dependent on the number of points he / she gained in physiology test. In order to increase internal accuracy (also referred to as curriculum or content accuracy) of both exams in the future (reaching value $r^2 > 0.5$), a precise plan of tests should be elaborated that would be consistent with the learning outcomes determined in the standards of teaching, including also cross-subject correlates [39].

Considering the above limitations, it may be assumed that to analyse the quality of recruitment criteria for candidates for medical studies at WUM, a good endpoint would be to carry out assessment from two exam subjects concerned with scientific basis of medicine: physiology and pathophysiology. The analysis of correlation showed that a criterion that functions very well is a sum of ranking points (r-Pearson in a range of 0.22 – 0.42), whereas partial subject criteria: biology and physics correlate very poorly with the results of pre-clinical exams ($r < 0.28$). Only chemistry, as an individual criterion, may be comparable in terms of the power of correlation with the sum of ranking points ($r < 0.35$). It could thus be concluded that using just a complex selection tool that assesses several fields of knowledge and skills is appropriate for accurate choice of good candidates.

Most validation data concerning entrance exams to medical schools come from correlative study on American MCAT [17-25]. The authors from the early 80s state that the power of correlation between the MCAT result and the students' achievements in their first years of studies ranges between 0.1 and 0.5, and on average it was around 0.3 [17-25]. Donnet *et al.* (2007) [23] set similar conclusions in his meta-analysis, when he claimed that 'The predictive validity of the MCAT ranges from small to medium for both medical school performance and medical board licensing exam measures.' The range of correlation coefficient values was between 0.39 and 0.60 [23]. Even stronger correlations were observed by the Association of American Medical Colleges (AAMC): 0.28-0.81 [24]. Also Julian *et al.* (2005) [25], in his cross-sectional study in which 14 universities took part, pointed out that the MCAT results and students' achievements in the early years of studies are correlated on the level of $r = 0.55$. Less numerous reports concern other exams, e.g. GAMSAT [8, 26]. Coates (2008) [8], confirmed the relevance of correlation of the GAMSTAT results with the students' achievements in their first years of medical studies. Also Groves *et al.* [26] (2007) demonstrated the relationship between the results of the third part of GASTAT (Reasoning in Biological and Physical Sciences) and scores in students of their sophomore year.

Apart from evaluating the results of entrance exam as an individual criterion of selection, the researchers undertook attempts of analysis of influence of other frequently used discrimination factor – GPA. In the meta-analysis carried out by Ferguson *et al.* (2002) [9] it was shown that the predictive value of the American MCAT in relation to future achievements of a student is significantly higher if GPA is an additional criterion [9]. However, there are contradictory opinions as for the usefulness of GPA as a good discrimination criterion, because in comparison with a student's achievements at pre-university education level, measured by GPA value, it is aptitude tests that are characterised by a lower influence of socio-economic factors [4]. On the other hand, despite the recommendation of the Admissions to Higher Education Steering Group which point to the need of using criteria of assessment from the cognitive and non-cognitive field while selecting candidates for higher education [27, 28], yet in many countries, including Poland, the only applied mechanism in the selection of candidates is an aptitude test conducted externally. It needs to be emphasised that he supporters of the purely cognitive system of recruitment draw our attention to the fact that applying complex and time-consuming systems of assessment of candidates in terms of them representing the non-cognitive

features desired in a doctor, is of very low predictive value in relation to achieving success in the future. Furthermore, the critics of such recruitment solutions point to the fact that it is rather the value of educational process within the medical studies that is of primary significance and predictive value than initial selection based on non-cognitive predispositions of a candidate [2, 40, 41].

5 CONCLUSIONS

Accuracy of selection of suitable candidates for medical studies ought to be evaluated on several levels as well as when monitoring the graduate's medical career. The aim should be to ensure that the appropriate recruitment process give a possibility to select the best candidates for future physicians who will be able to deal with the science of complex both pre-clinical and clinical competences.

REFERENCES

[1] *The Central Committee of Examination*. [cited 18-05-2014] Available from: http://www.cke.edu.pl/

[2] McManus, I.C., et al. (2005).*Unhappiness and dissatisfaction in doctors cannot be predicted by selectors from medical school application forms: a prospective, longitudinal study.*BMC Medical Education5, pp. 38.

[3] Gwarek, K. (2008). *Ruszyła rekrutacja.* Medycyna Dydaktyka Wychowanie 40(5), pp. 32.

[4] Cerutti, B., Bernheim, L., van Gessel, E. (2013).*The predictive validity of the aptitude test for the performance of students starting a medical curriculum.*Swiss Medical Weekly 143, pp. w13872.

[5] Głowacka, M.D., Pagórski, P. (2011).*Budowa ścieżek karier pracowniczych w zakładzie opieki zdrowotnej*, in *Zarządzanie zakładem opieki zdrowotnej: Konteksty i determinanty zmian*, M.D. Głowacka, Editor. Polskie Towarzystwo Nauk o Zdrowiu: Poznań.

[6] *Regulation of the Minister of Science and Higher Education on education standards for courses of study: medical, dental, pharmacy, nursing and midwifery*. May 9, 2012, Ministry of Science and Higher Education: Dz.U. 2012 nr 0 poz. 631.

[7] Maggio, L.A., et al. (2013).*Evidence-based medicine training in undergraduate medical education: a review and critique of the literature published 2006-2011.* Academic Medicine 88(7), pp. 1022-8.

[8] Coates, H. (2008).*Establishing the criterion validity of the Graduate Medical School Admissions Test (GAMSAT).* Medical Education 42(10), pp. 999-1006.

[9] Ferguson, E., James, D., Madeley, L. (2002). *Factors associated with success in medical school: systematic review of the literature.* BMJ 324(7343), pp. 952-7.

[10] Al Alwan, I., et al. (2013). *Health sciences and medical college preadmission criteria and prediction of in-course academic performance: a longitudinal cohort study.*Advances in Health Sciences Education: Theory and Practice 18(3), pp. 427-38.

[11] Reiter, H.I., et al. (2007). *Multiple mini-interviews predict clerkship and licensing examination performance.*Medical Education 41(4), pp. 378-84.

[12] Mercer, A., Puddey, I.B. (2011). *Admission selection criteria as predictors of outcomes in an undergraduate medical course: a prospective study.*Medical Teacher 33(12), pp. 997-1004.

[13] Ronai, A.K., et al. (1984). *Relationship between past academic performance and results of specialty in-training examinations.* Journal of Medical Education 59(4),pp. 341-4.

[14] Mennin, S.P., et al. (1993). *Performances on the NBME I, II, and III by medical students in the problem-based learning and conventional tracks at the University of New Mexico.* Academic Medicine 68(8), pp. 616-24.

[15] Richards, J.M., Taylor, C.W.,Price, P.B. (1962). *The prediction of medical intern performance.* Journal of Applied Psychology 46(2), pp. 142-146.

[16] Glaser, K., et al. (1992). *Science, Verbal, or Quantitative Skills: Which is the Most Important Predictor of Physician Competence?* Educational and Psychological Measurement 52(2), pp. 395-406.

[17] Brooks, C.M., et al. (1981). *Validity of the new MCAT for predicting GPA and NBME part I examination performance.* Journal of Medical Education 56(9 Pt 1), pp. 767-9.

[18] Carline, J.D., et al. (1983). *Predicting performance during clinical years from the new Medical College Admission Test.* Journal of Medical Education 58(1), pp. 18-25.

[19] McGuire, F.L. (1982). *The new MCAT and its relationship to medical student performance-year two.* Journal of Medical Education57(1), pp. 60-1.

[20] McGuire, F.L. (1980). *The new MCAT and medical student performance.* Journal of Medical Education55(5), pp. 405-8.

[21] Hynes, K., Givner, N. (1981). *Restriction of range effects on the new MCAT's predictive validity.* Journal of Medical Education 56(4), pp. 352-3.

[22] Cullen, T.J., et al. (1980). *Predicting first-quarter test scores from the new Medical College Admission Test.* Journal of Medical Education 55(5), pp. 393-8.

[23] Donnon, T., Paolucci, E.O.,Violato, C. (2007). *The predictive validity of the MCAT for medical school performance and medical board licensing examinations: a meta-analysis of the published research.*Academic Medicine 82(1), pp. 100-6.

[24] Association of American Medical Colleges. (2000). *The predictive validity of the Medical College Admission Test.* Contemporary Issues in Medical Education 3(2), pp. 1-2.

[25] Julian, E.R. (2005). *Validity of the Medical College Admission Test for predicting medical school performance.*Academic Medicine 80(10), pp. 910-7.

[26] Groves, M.A., Gordon, J., Ryan G. (2007). *Entry tests for graduate medical programs: is it time to re-think?* The Medical Journal of Australia 186(3),pp. 120-3.

[27] Schwartz, S. (2004). *Fair admissions to higher education: recommendations for good practice*. Department for Education and Skills Publications.

[28] Roberts, C., Prideaux, D. (2010). *Selection for medical schools: re-imaging as an international discourse.*Medical Education 44(11), pp. 1054-6.

[29] Feldt, L.S. (1969). *A test of hypothesis that Cronbach's alpha or Kuder-Richardson coefficent 20 is same for 2 tests.* Psychometrika 34(3), pp. 363.

[30] Kaplan, R.M., Satterfield, J.M., Kington, R.S. (2012). *Building a better physician--the case for the new MCAT.* The New England Journal of Medicine 366(14), pp. 1265-8.

[31] Adam, J., Dowell, J., Greatrix, R. (2011). *Use of UKCAT scores in student selection by U.K. medical schools, 2006-2010.* BMC Medical Education 11, pp. 98.

[32] Callahan, C.A., et al. (2010). *The predictive validity of three versions of the MCAT in relation to performance in medical school, residency, and licensing examinations: a longitudinal study of 36 classes of Jefferson Medical College.*Academic Medicine 85(6), pp. 980-7.

[33] Prideaux, D., et al. (2011).*Assessment for selection for the health care professions and specialty training: consensus statement and recommendations from the Ottawa 2010 Conference.* Medical Teacher 33(3), pp. 215-23.

[34] Reznick, R.K., et al. (1993). *An objective structured clinical examination for the licentiate of the Medical Council of Canada: from research to reality.*Academic Medicine 68(10 Suppl), pp. S4-6.

[35] McManus, I.C., et al. (2013).*The Academic Backbone: longitudinal continuities in educational achievement from secondary school and medical school to MRCP(UK) and the specialist register in UK medical students and doctors.*BMC Medicine 11, pp. 242.

[36] Julian, E.R. (2005).*Validity of the Medical College Admission Test for predicting medical school performance.*Academic Medicine 80(10), pp. 910-7.

[37] Guilford, J.P. (1954). *Psychometric methods*. 2nd ed. 1954, New York: McGraw-Hill.

[38] Goodwin, L.D. (2002). *Changing conceptions of measurement validity: an update on the new standards.* The Journal of Nursing Education 41(3), pp. 100-6.

[39] Ebel, R.L. (1961). *Must all tests be valid?* American Psychologist 16(10), pp. 640-647.

[40] Norman, G. (2004). *The morality of medical school admissions.*Advances in Health Sciences Education: Theory and Practice 9(2), pp. 79-82.

[41] Peskun, C., Detsky, A., Shandling, M. (2007).*Effectiveness of medical school admissions criteria in predicting residency ranking four years later.*Medical Education 41(1), pp. 57-64.

ANNEX

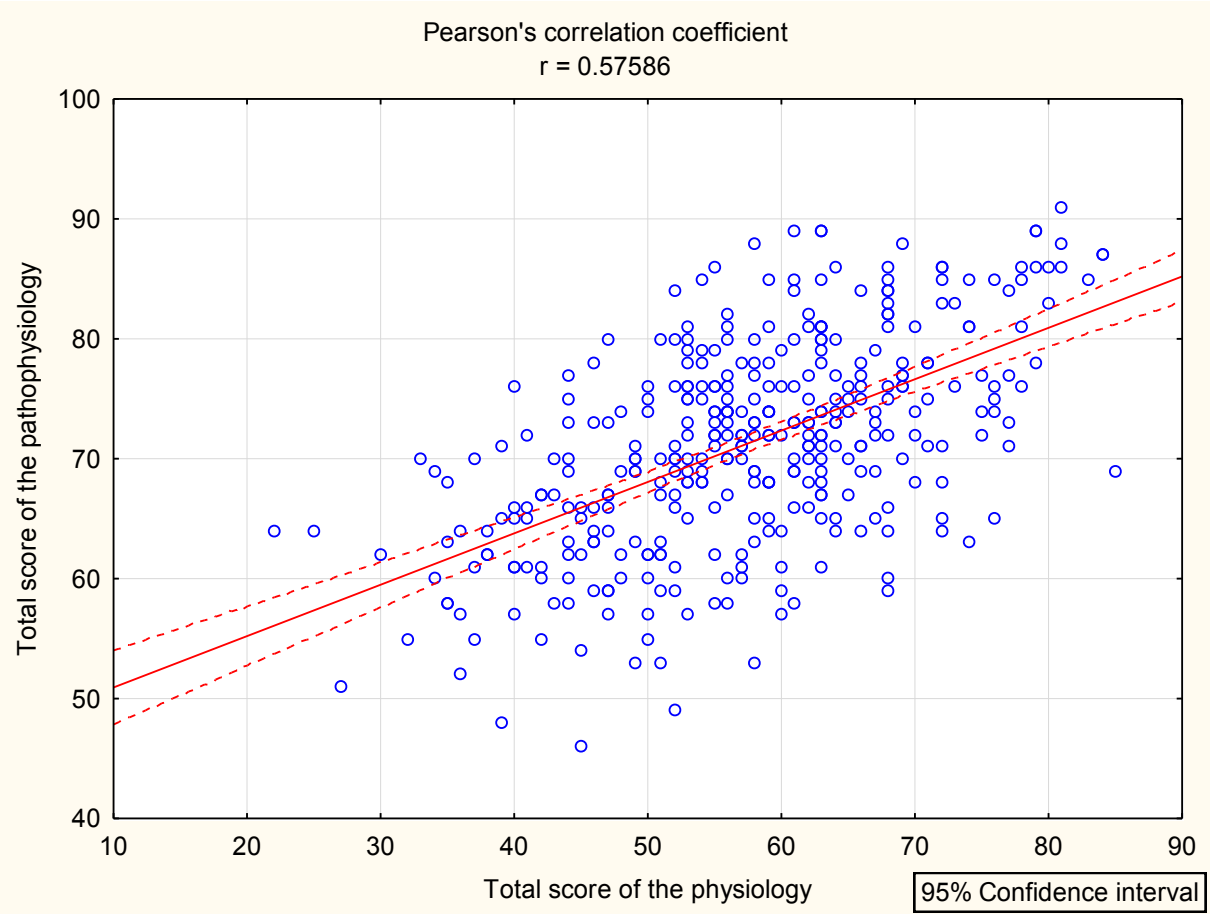

Figure 1. Example of a graph with correlation between a score in the exam in physiology and pathophysiology doe the 1st Medical Faculty between the years 2009-2010.

Table 1. Summary of quality parametres of the subject exam results in physiology and pathophysiology for two years of students from the first and the second year of Medical Faculty at WUM.

	Physiology 2009 (I i II WL)	Physiology 2011 (I i II WL)	Pathophysiology 2010 (I WL)	Pathophysiology 2010 (II WL)	Pathophysiology 2012 (I WL)	Pathophysiology 2012 (II WL)
Number of questions	100	100	99	100	98	100
Number of options in a test question	4	4	5	5	5	5
Number of versions	2	2	2	2	4	4
Number of cases	416	446	280	102	325	112
*Normal distribution**	*P* = 0.048	*P*< 0.001	*P*> 0.05	*P* = 0.006	*P*< 0.0001	*P*> 0.05
*Data outliers***	*P*> 0.05	*P* = 0.01	*P*> 0.05	*P*> 0.05	*P*> 0.05	*P*> 0.05
average	71.5	80.0	54.2	67.9	69.3	58.9
Median	72	81	55	69	70	59
Minimum	35	38	22	40	37	32
Maximum	93	98	84	86	89	86
Lower quartile	65	74	47	62	63	51
Upper quartile	78	88	62	75	76	68
Standard deviation	9.40	10.24	10.58	9.83	8.98	11.75
Obliqueness	-0.312	-0.708	-0.199	-0.723	-0.571	0.064
Cronbach's α	0.825	0.884	0.834	0.841	0.810	0.874
Number of questions with negative correlation	7	0	9	13	5	8
Number of question with negative differentiating power	1	1	2	6	2	3
Index of test difficulty	0.71	0.80	0.55	0.68	0.70	0.60
Cronbach's α *after optimization*	0.843	0.884	0.851	0.871	0.824	0.887

***** Shapiro-Wilk test (for *P*< 0.05 distribution is not normal).
** Grubbs test for outliers (for *P*< 0.05 presence of outliers).

Table 2. Matrix of correlation of the sum of points gained in subject exams of physiology and pathophysiology and evaluation of reliability of pre-clinical competences in these two thematic areas.

	Physiology 2009 (I WL)	**Physiology 2009 (II WL)**	**Physiology 2011 (I WL)**	**Physiology 2011 (II WL)**
Pathophysiology 2010 (I WL)	r = 0.624* α = 0.882			
Pathophysiology 2010 (II WL)		r = 0.530* α = 0.872		
Pathophysiology 2012 (I WL)			r = 0.521* α = 0.897	
Pathophysiology 2012 (II WL)				r = 0.589* α = 0.922

*P< 0.001
Cronbach's α

Table 3. The analysis of correlation of subject exam results with the entrance exam results in biology, chemistry and physics for two years of students, divided according to the form of the undertaken medical studies.

Subject exam	**Sum of the recruitment criteria**	**criterion: BIOLOGY**	**criterion: CHEMISTRY**	**criterion: PHYSICS**
Recruitment 2008				
Physiology (2009)				
- all students	r = 0.368*	r = 0.221*	r = 0.345*	r = 0.247*
- full-time	r = 0.259*	r = 0.072†	r = 0.230*	r = 0.143***
- part-time	r = 0.343**	r = 0.173†	r = 0.290***	r = 0.013†
Pathophysiology (2010)				
- all students	r = 0.417*	r = 0.272*	r = 0.349*	r = 0.276*
- stacjonarne	r = 0.259*	r = 0.113†	r = 0.183**	r = 0.126***
- niestacjonarne	r = 0.363**	r = 0.124†	r = 0.271***	r = 0.110†
Recruitment 2010				
Physiology (2011)				
- all students	r = 0.217*	r = 0.134**	r = 0.127**	r = 0.111***
- full-time	r = 0.232*	r = 0.123***	r = 0.127**	r = 0.091†
- part-time	r = 0.045†	r = -0.72†	r = 0.043†	r = 0.035†
Pathophysiology (2012)				
- all students	r = 0.256*	r = 0.148**	r = 0.117***	r = 0.122***
- full-time	r = 0.316*	r = 0.151**	r = 0.113***	r = 0.091†
- part-time	r = 0.069†	r = -0.124†	r = 0.050†	r = 0.171†

r – Pearson linear correlation coefficient
*P< 0.001, ** P< 0.01, *** P< 0.05, $^{\dagger}P$> 0.05