

Good practice in quality assurance of validation

United Kingdom

Qualification "Physician"



Good practise in quality assurance of validation. Great Britain. Qualification „Physician”

Based on the expertise by Katarzyna Ludka-Stempień

Editors:

Sylvia Walicka, Iwona Gmaj, Barbara Przybylska

Instytut Badań Edukacyjnych

ul. Górczewska 8

01-180 Warszawa

tel. (22) 241 71 00; www.ibe.edu.pl

Citation template:

Educational Research Institute (2020). *Good practise in quality assurance of validation. Great Britain. Qualification „Physician”*. Warsaw: Educational Research Institute.

© Copyright by: Instytut Badań Edukacyjnych, Warsaw 2020

ISBN 978-83-66612-47-1

This publication is co-financed by the European Social Fund of the European Union as part of the project entitled *Support to central government administration, awarding bodies and quality assurance institutions in implementing stage II of the Integrated Qualifications System*.

This copy is free of charge.

Table of contents

1. Country, institution name.	4
2. Institution website.	4
3. Qualifications.	4
4. Good practice - External quality assurance.	4
4.1. Background Information	6
4.2. Current Form of the MRCP(UK) Examination	6
4.2.1 Part i of the Examination	7
4.2.2 Part II of the Examination.	10
4.2.3. Practical Assessment of Clinical Examination Skills (PACES)	12
4.2.4 Reliability and Accuracy of Examination Results	16
4.3. Centuries of Experience Leading to MRCP(UK).	16
4.3.1. The Royal College of Physicians of London	16
4.3.2. The Act of 1858 and its consequences	17
4.3.3 Beginning of the 20th century	18
4.3.4. The 1960s and the increase in recruitment to higher education	19
4.3.5. From the 1970s to the present	20
4.3.5.1. Format of questions and marking	20
4.3.5.2. Psychometric evaluation.	20
4.3.5.3. Separation of PACES and assessing its accuracy and reliability	21
5. Comments.	23

1. Country, institution name

- United Kingdom
- The Royal College of Physicians of London (RCP)

2. Institution website

<https://www.rcplondon.ac.uk/>

3. Qualifications

Examination for Membership of the Royal Colleges of Physicians of the United Kingdom: MRCP(UK).

4. Good practice - External quality assurance

Improvement of the assessment process: the MRCP(UK) medical qualification

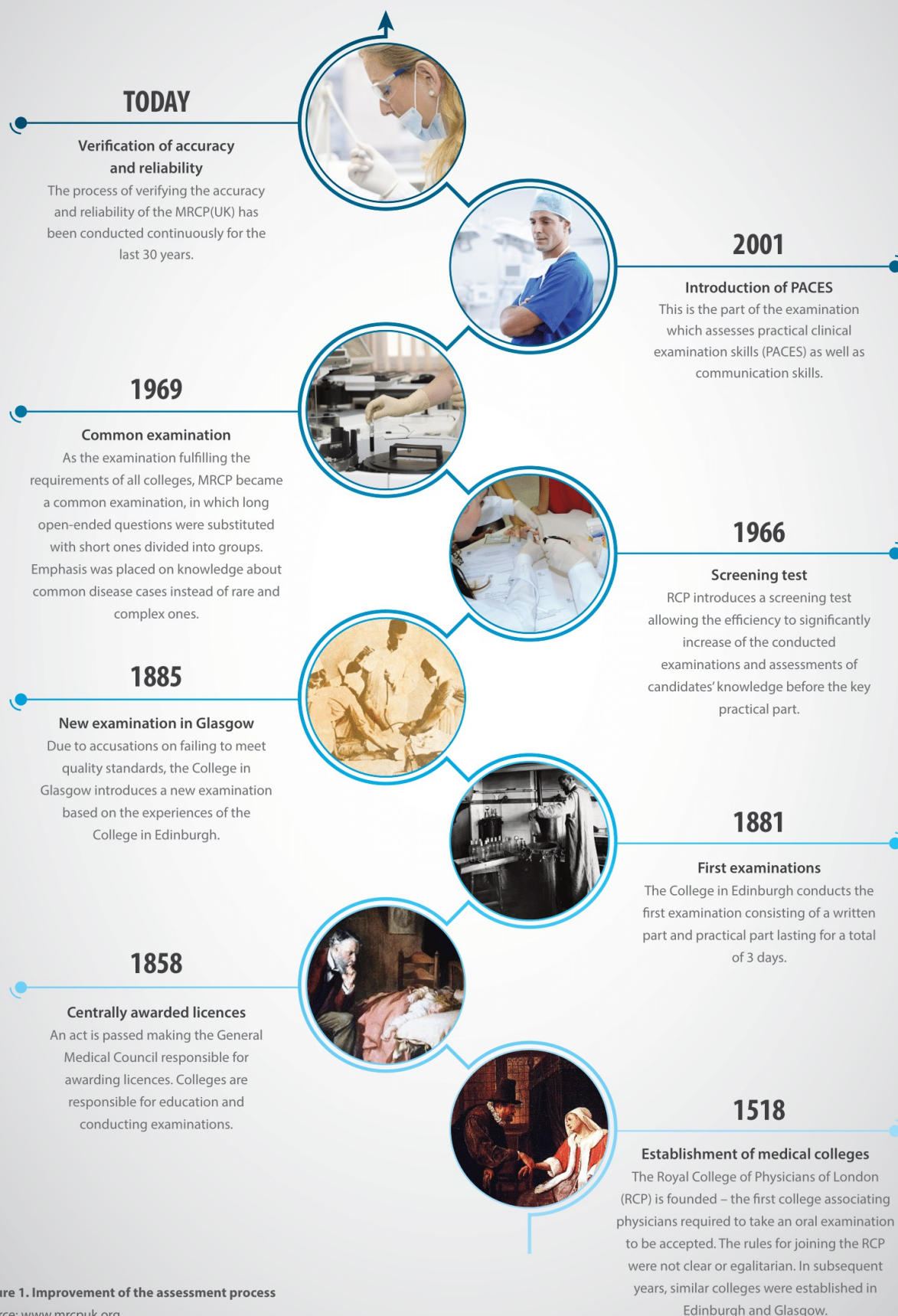


Figure 1. Improvement of the assessment process
Source: www.mrcpuk.org

4.1. Background Information

The good practice refers to the quality assurance of the MRCP(UK) examination, i.e. the compulsory examination required for the medical career path in Great Britain, organised by three medical colleges in London, Edinburgh and Glasgow (Royal Colleges of Physicians, RCP). The colleges are institutions which, pursuant to a royal edict or a legal act passed by the Parliament, received the right to control the teaching programme in a specified area of knowledge. Therefore, they perform the role of professional associations.

The full name of the examination is: the Examination for Membership of the Royal Colleges of Physicians of the United Kingdom ? MRCP(UK). The English term physician means an internal medicine specialist.

The MRCP(UK) is an international medical examination organised on a broad scale. Every year, it is taken by approx. 6,000 persons from all over the world, whereas only 30% are graduates of British schools. Apart from Great Britain, the examination is held in 20 other countries, including India, the United Arab Emirates, Egypt and Sri Lanka. The MRCP(UK) is not only an examination of international renown, but also a good example of a reliable and accurate measuring tool, the history of which shows the way in which social changes and the development of knowledge affect the process of validation and awarding qualifications.

The good practice refers to two aspects:

- currently functioning solutions within the scope of ensuring the quality of the validation of the qualification of physician
- improving the description of this qualification, recorded for several centuries, and the process of validation leading to its formal confirmation.

Please note that solutions pertaining to the description of the MRCP(UK) qualification and the examination leading to its confirmation are continually improved, whereas the presented good practice contains a description of the solutions as of the end of 2016.

4.2. Current Form of the MRCP(UK) Examination

The MRCP(UK) examination has been devised to enable the selection of the best candidates for physicians, who present adequate knowledge, skills and behaviours which, in the college's evaluation, allow for completion of a specialty. Passing the MRCP(UK) examination guarantees access to the continuation of education in the following specialties: dermatology, endocrinology, gastroenterology, geriatrics, infectious diseases, oncology, nephrology, neurology, palliative medicine, pulmonology and rheumatology. The Royal Colleges of Physicians also organise, in cooperation with the British Cardiovascular Society, a specialty in cardiology. The MRCP(UK) examination is also required by some specialist colleges, e.g. the Royal College of Radiologists.

The examination consists of three components.

- part I: written multiple-choice test
- part II: written examination
- part III: Practical Assessment of Clinical Examination Skills (PACES)

The examination takes place over three sessions (corresponding to the individual parts) organised in **various months of the year**. Before 2009, candidates took the individual parts of the examination in line with the sequence specified above, whereas now, it is also possible to take part III (PACES) first and then take part II. Part I is to be taken at the very beginning. Up-to-date information about the examination, sample questions, principles of conduct, advice and all additional information for candidates are available at the website: www.mrcpuk.org. The candidates receive the results of the examination in electronic form after logging in to the system.

The candidates for physicians usually try to pass the first part of the examination during the second year of their study programme, yet this is not a formal requirement of the RCP. The regulations stipulate that part i may be taken up to 12 months after completion of education at university or after the end of the first year of the basic programme (RCP, 2016d). In line with the regulations, the candidate has a maximum of six attempts at passing every part and has seven years to pass the entire examination. After this deadline, all the parts have to be passed again. The candidates who do not pass the examination (have used their 6 chances) must choose another medical specialty.

Receipt of the MRCP(UK) qualification is a time-consuming process. Data about the passing rate of candidates between 2003 and 2011 show that approx. 28% of persons who take the examination receive the MRCP(UK) diploma, whereas candidates need two attempts on average to pass part I, 1.6 attempts to pass part II and 1.8 attempts to pass PACES (Ludka-Stempień, 2015). Only 8% of all candidates taking the examination pass all the parts in the first attempt.

Receipt of the MRCP(UK) qualification requires significant financial outlays by the candidate. Taking all the parts of the examination costs a total of approx. GBP 1,500, with the assumption that the candidate passes each part only once, which rarely happens. The fee for the examination only covers the examination costs. Part i and II are conducted in examination centres; the candidates answer questions on paper test sheets; the result is calculated digitally after the sheets have been scanned. The practical examination takes place in hospital wards.

4.2.1 Part i of the Examination

The purpose of part i is to evaluate the candidate's knowledge and understanding of the exact sciences relating to medicine and their knowledge and experience of common or important diseases to a degree that allows them to begin specialist training. The examination consists of two three-hour tests, which take place during one day. Each test has one hundred multiple-choice questions, where only one out of five answers is correct. Questions refer to 14 areas of general medicine in the proportions described in Table 1.

The test questions are derived from a general pool. Each question can be used only five times (irrespective of the time interval), and after the fifth time, it cannot be included in an examination for at least three years.

The examination committee consists of 18 elected members of the college. Before every session, the committee reviews the questions and analyses them in detail with respect to language, scope, difficulty, discriminative power and quality of multiple choice answers. Questions that arouse doubts are substituted with other questions from the pool. In this manner, the final content of the test is determined, which is approved by the entire committee.

Area	Number of questions
Cardiology	15
Haematology and oncology	15
Pharmacology, therapy, toxicology	16

Exact and medical sciences:	25
Cell biology	2
Anatomy	3
Metabolism and organism bio-chemistry	4
Physiology	4
Genetics	3
Immunology	4
Statistics, epidemiology and medicine based on scientific evidence	5
Dermatology	8
Endocrinology	15
Geriatrics	4
Gastroenterology	15
Infectious diseases	15
Neurology	15
Nephrology	15
Ophthalmology	4
Psychiatry	8

Pulmonology	15
Rheumatology	15
Total	200

Table 1. Areas of the questions for part i of the examination

Source: Ludka-Stempień (2015), quoted from www.mrcpuk.org

Not all the answers given by the candidate are included in the test assessment. Approx. 10% of 200 questions are not assessed; their purpose is the statistical levelling of results among individual sessions. Questions considered defective are also not taken into account during the final assessment.

Results of part i and II of the examination are announced four weeks after the examination date. Figure 3 presents the process of the assessment encompassing the calculation of results and the designation of the mark threshold required to pass the examination. This is a multi-stage process, which assumes a renewed verification of the quality of the test items, removal of defective items and leaving only those that will be marked. After the calculations, the results are presented on a scale from 0 to 99 points.

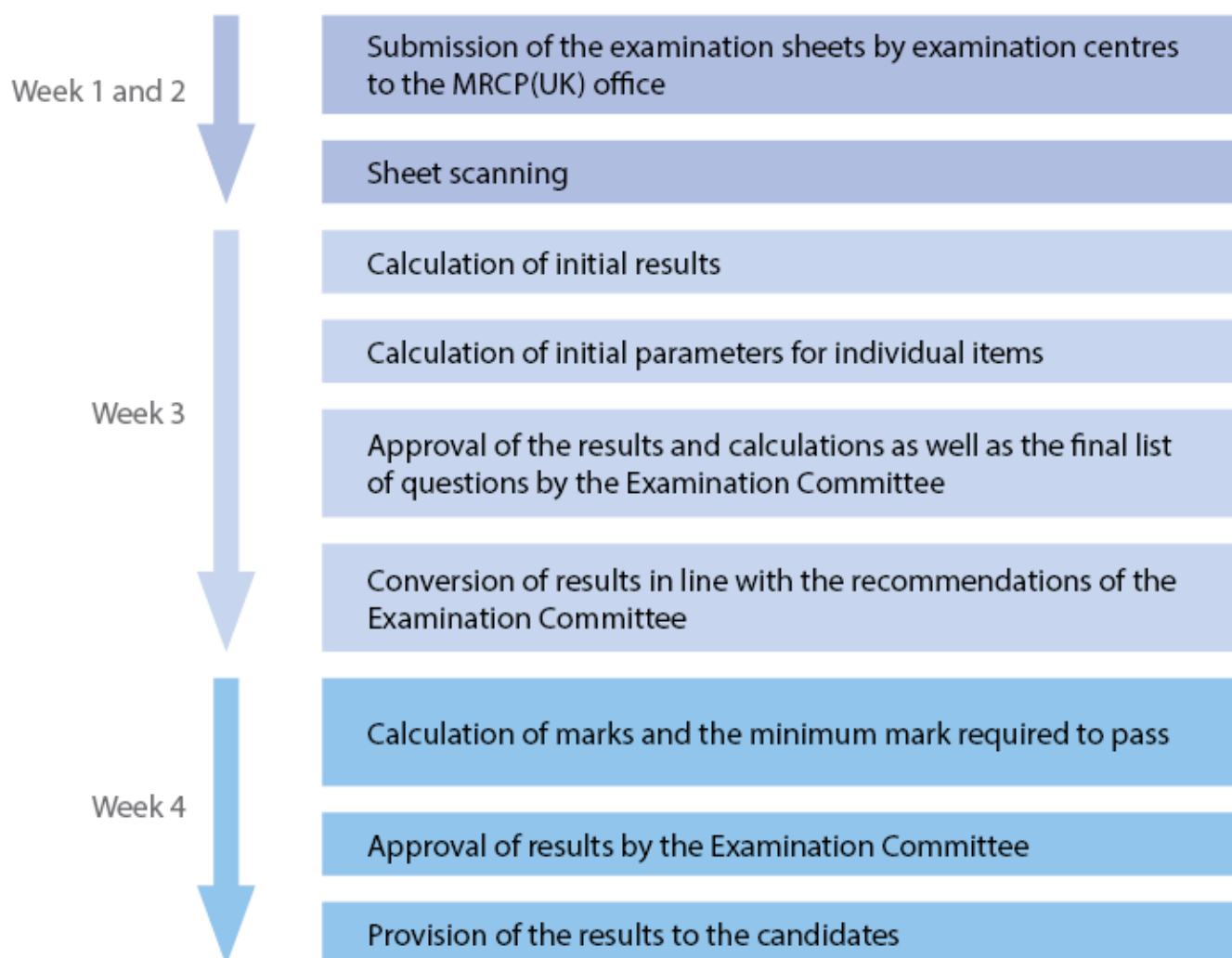


Figure 2. Process of calculating the results of parts I and II of MRCP(UK)

Source: www.mrcpuk.org

4.2.2 Part II of the Examination

Part II is a written test aimed at verifying whether the candidate is capable of applying medical knowledge to specific patient cases, but not yet under simulated conditions as in the practical part of the examination (PACES). The questions refer to, among others, analyses of patients' test results, offering a diagnosis and prognosis. This part consists of three written tests taken by the candidate over two consecutive days. Each test has 90 questions in a format similar to part I. Questions often contain a description of a clinical case, results of medical tests, and may also feature illustrations or photographs. Sample questions are presented on the examination's website. As was the case in part I, the examination committee also chooses the questions for part II, relying on psychometric data from strictly defined areas, in the proportions reflected in Table 2.

As with part I, every question can only be used five times (irrespective of the time interval) and then a three-year break is required before it can be used again. Part II of the examination also contains questions which are not given a mark and used to compare the level of difficulty among the examinations over the years.

Area	Total number of questions
Cardiology	25
Dermatology	13
Endocrinology and metabolism	25
Gastroenterology	25
Geriatrics	10
Haematology	13
Infectious diseases	25
Neurology	25
Nephrology	25
Oncology and palliative medicine	13
Ophthalmology	4

Psychiatry	4
Pulmonology	25
Rheumatology	13
Therapeutics and toxicology	25
Total	270

Table 2. Areas of the questions and their number for part II of the examination

Source: Ludka-Stempień (2015), quoted from www.mrcpuk.org

The process of assessing the examination sheets is analogous to the one applied in part I. Examinations take place in designated examination centres, whereas the calculation of the parameters of individual examination items, individual results and the threshold required to pass the examination takes four weeks. After this period, the candidates receive information about their results.

4.2.3. Practical Assessment of Clinical Examination Skills (PACES)

PACES verifies whether the candidate has achieved the specific competences required in clinical practice, such as:

- patient interviewing
- communicating results to patients, their families and colleagues
- performing tests and interpreting symptoms correctly
- conducting proper diagnoses
- being able to plan and discuss immediate actions and long-term treatment
- taking ethical issues into account in line with the criteria specified in *Good Medical Practice* (GMC, 2013a).

Presently, the examination lasts two hours and is conducted at “clinical stations”. At each station, candidates are given tasks aimed at assessing their skills and knowledge representing various areas of internal medicine and communication skills. At each station, different cases of patients are described with problems pertaining to: (i) respiratory and abdominal problems (two cases at 10 minutes each), (ii) neurological and cardiovascular problems (two cases at 10 minutes each), (iii) skin diseases, balance disorders, eye diseases and hormonal problems (four cases at five minutes each) - three stations in total. At the last two stations, the candidates must interview the patient, during which their communication skills and professional ethics are assessed. 20 minutes are assigned for the tasks performed at each station, which corresponds to the work performed at a hospital, where this is the amount of time a physician has to evaluate the condition of a patient. Five-minute breaks are scheduled between the tasks. The beginning and end of the task at each station is announced by a bell. Before starting a task, the candidate prepares during the breaks and has five minutes before the start of the first task for this purpose. During this time, the candidate also reads the task scenario,

which may contain: the patient's results, introduction to the interview with the patient, instructions. The examination is conducted simultaneously for five candidates, each of whom solves tasks at a different station. After the task is performed, the candidates move to different stations.

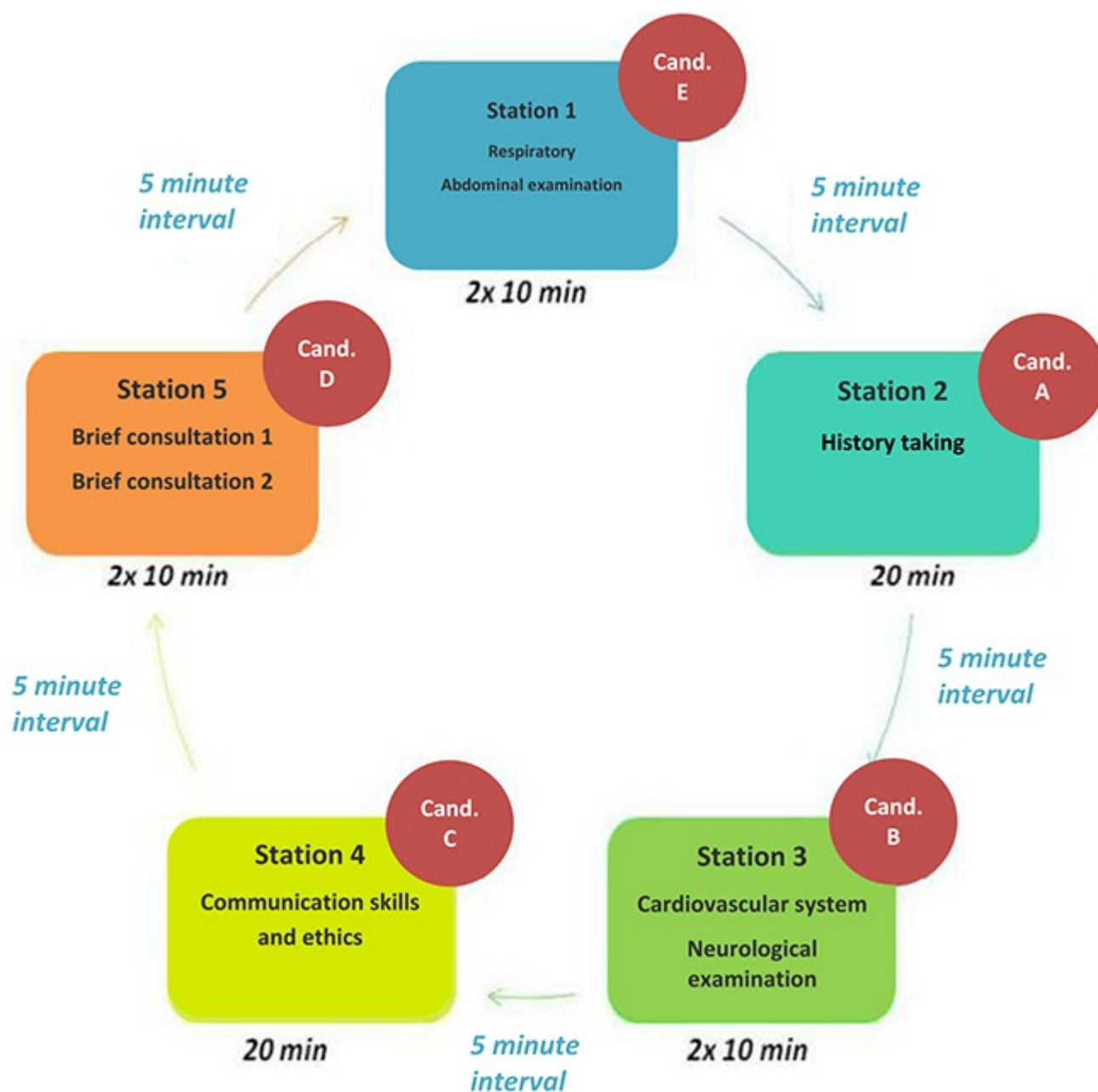


Figure 3. Carousel of PACES stations

Source: Ludka-Stempień (2015), quoted from www.mrcpuk.org

During the examination, the candidates analyse cases of authentic persons or professional actors play the role of patients. They mainly appear in tasks pertaining to history taking and relate to communication and professional ethics. Detailed guidelines pertaining to the way in which the examination should be conducted have been prepared. The guidelines also include such elements as, e.g., the arrangement of furniture in the examination room (RCP, 2016c).

Skill	Abdominal examination	Respiratory examination	History taking	Cardiovascular examination	Neurological examination	Communication and ethics	Brief consultation 1	Brief consultation 2	Min. no. of marks to be received
A. Physical examin.	X	X		X	X		X	X	14
B. Identifying physical signs	X	X		X	X		X	X	14
C. Clinical communication			X			X	X	X	10
D. Differential diagnosis	X	X	X	X	X		X	X	16
E. Clinical judgement	X	X	X	X	X	X	X	X	18
F. Manag. patient's concerns			X			X	X	X	10
G. Maintaining patient welfare	X	X	X	X	X	X	X	X	28

Table 3. List of skills assessed during PACES

Source: Ludka-Stempień (2015), quoted from www.mrcpuk.org

The candidate must receive a minimum number of points from each set of skills contained in Table 3. The Table also specifies the task in which a given skill is assessed. In total, the candidates must receive 130 points. This is a higher value than the total number of minimum points (110 points), which means that passing the examination requires achieving more points than the minimum in some sets of skills.

Two examiners independently assess the tasks performed at each station. One examiner talks to the candidate and asks questions. The other analyses the course of the session. The examiners swap roles during the assessment of subsequent candidates. The examiners cannot comment on the course of the examination and should not answer candidates' questions or prompt any questions. During the examination, the examiners cannot express their opinion about the candidate in any manner or about what the candidate says or does. These procedures are intended to help maintain proper standards for conducting the examination. The examiners undergo detailed training and receive specific written guidelines (RCP 2016c).

A three-level marking scale is used:

- satisfactory
- borderline
- unsatisfactory

The examiners who decide to give a low or borderline mark to a candidate are required to justify their decision in writing, due to the fact that receiving such marks results in failing the examination. Furthermore the examiners, if they deem it necessary, may provide the candidate with feedback.

A PACES examiner must fulfil the following criteria:

» if such person is a physician in Great Britain:

- have the MRCP (UK) diploma
- be a fellow or a collegiate member of one of the three Royal Colleges of Physicians (college members are required to have a certification of completion of specialty training for at least two years and to work in the position of a consultant)
- be registered with the General Medical Council (GMC)
- participate in in-service training, undergo an annual evaluation at his/her work place and show evidence of acquiring new competences

» if such person is a physician outside of Great Britain:

- they must be a fellow in one of the three Royal Colleges of Physicians
- they must be entered in the register of physicians and licensed to practice the profession

» and furthermore (joint criteria for all physicians), the person:

- has practised medicine in the course of the last two years at a hospital or other medical establishment
- evaluates, examines and handles clinical cases on a daily basis within the spectrum of issues covered by PACES
- supervises the training of physicians at the level of the main specialty training (CMT) or an analogous one, who takes care of patients with illnesses from the spectrum of the issues covered by PACES
- completed training on equality and diversity within the last three years
- commits to participating in at least six examination cycles (30 candidates) per year (averaged over 2 years), which is a minimum of 2-3 days examining per year

The examiners complete special training to prepare them to be consistent in their marking when assessing candidates and to maintain the reliability of the examination. This training takes place during PACES examination sessions. First, applicant examiners participate in a pre-examination meeting, during which the actual examiners determine which physical symptoms should be assessed by the candidates, which

questions may be asked by the examiners, and how the course of the examination will be assessed. Next, an applicant examiner joins one of the stations as an observer and assesses the candidate's skills as do the actual examiners. The mark sheet completed by the applicant examiner is later compared with the actual examiners' sheets and discussed. After three examination cycles (i.e. at the end of the day), the applicant examiner's sheets are evaluated and he/she is given feedback from the actual examiners. If the applicant examiner still wants to be an examiner and was well evaluated, he/she will now be able to independently assess candidates during PACES (RCP, 2016b).

4.2.4 Reliability and Accuracy of Examination Results

The reliability indicators of all parts of the examination are at a high level, i.e. above 0.90 for parts I and II and 0.80 for PACES. Accuracy is confirmed primarily by testing the significance of the statistical relationships between the results of individual parts of the MRCP(UK) and the results of 17 other written examinations and 2 clinical examinations in other medical specialties, as well as in relation to annual evaluations and cases of revoked licences or limitations placed on the right to practice the profession by the GMC. These correlations were stronger in the case of written examinations and weaker for the assessment of practical clinical skills (Ludka-Stempień, 2015). The key conclusion from these analyses is the fact that persons who receive low marks in the examination stand a greater risk of subsequent problems with the licence, decreased chances for good marks during training and decreased chances for good examination results at the end of training, which confirms the forecasting accuracy of the examination.

4.3. Centuries of Experience Leading to MRCP(UK)

The present-day formula of the examination is the result of many years of tradition of giving medical examinations with the assumption of their constant improvement. The changes introduced resulted from both the development of knowledge and technological progress, as well as social changes. The description of these changes explains why the MRCP(UK) certificate is considered prestigious.

4.3.1. The Royal College of Physicians of London

Establishing colleges as institutions formally awarding physician qualifications was the first attempt to ensure high quality medical services.

The Royal College of Physicians of London was the first college associating physicians. It was established in 1518 by King Henry VIII pursuant to an edict, which was issued in response to a petition filed by seven doctors providing services in the region of contemporary London. At that time, the services of doctors were reserved exclusively for the aristocracy. The rest of society used the services of herbalists, barber-surgeons, and often ordinary charlatans. The aforementioned petition to the king had two purposes: to ensure the quality of the services provided by eliminating persons lacking adequate competences, and also to limit the competition in medical services. The royal act provided the college with exclusivity in the provision of medical services within the city of London.

Shortly after the college was founded in London, the Royal College of Physical Surgeons of Glasgow was established in 1599. Much later, in 1681, the Royal College of Edinburgh was founded. Until 1968, the colleges functioned independently, had their own spheres of influence and, importantly, offered individual membership. This also meant that they had their own examinations, which were similar to a certain degree.

Entrance examinations to the London college still serve as an example of a typical contemporary medical examination. The London college offered two types of membership and two titles: licentiate and fellow. The two categories differed significantly with respect to rights: licentiates had the right to practice medicine, whereas fellows could additionally have a say about the established provisions that had to be observed by all college members. In addition, they had the right to participate in the college's authorities and vote in the elections of such authorities (Waddington, 1973).

Membership in the college was conferred upon referral, after a presentation of diplomas from renowned

universities or on the basis of evidence of operating a good practice outside of the college's region, e.g. in another country. The examination for a fellow was conducted orally and the mark was given by persons called "censors" (who can be considered the equivalents of examiners). This was followed by voting. No records are preserved with the detailed questions that the examiners asked. The principles of joining the college were considered, at any rate, unclear and definitely not egalitarian. It is known that full membership required at least four years of medical practice. In order to make this condition possible, the examination could be conditionally passed before the completion of the required practice, and subsequently, after four years, it had to be taken again. Such conditional members were called candidates.

The college had few members. In the 16th century, the highest recorded number was 31 (Clark, 1965). In 1708, only 57 persons had full member rights. Additionally, there were seven candidates and three fellows. In comparison, there were 39 licentiates. These numbers were even lower in 1746 (RCP and Cuthbertson, 2008). To change the proportions in the number of fellows to licentiates, due to the dissatisfaction of the latter with respect to the limitation of their rights, a new statute was introduced in 1771 enabling licentiates who practised for at least seven years and were older than 35 to be nominated as a candidate by any college member. In accordance with the regulations, such a candidate had to pass an examination (unfortunately, its scope was not specified) in medicine and be familiar with the works of such philosophers as Hippocrates, Galen and Aretheus. However, it was still believed that a candidate passed the examination if the college members confirmed this by voting. This formula was preserved until the mid-19th century.

4.3.2. The Act of 1858 and its consequences

The next steps in the quality assurance of medical qualifications were the establishment of the General Medical Council as an institution granting medical licences independently from colleges and an attempt made to standardise examinations.

At the beginning of the 19th century, London already had half a million residents. The hitherto number of physicians did not fulfil the needs of such a large population and both this situation, as well as the interests of college members put pressure on the London college to increase the number of licensed physicians (Waddington, 1973). The Victorian era was also a period of the development of medical knowledge, the popularisation of the idea of professionalization, but additionally of a growing concern for public health. In this context, the adoption of the Medical Act in 1858 was of great significance (Roberts, 2009).

The act was received by the medical community with moderate enthusiasm. So far, colleges had absolute power over the process of granting licences, due to the fact that receiving one was tantamount to membership in the college. As of the moment of Parliament's adoption of the act, the authority of colleges in this respect became limited and they lost their spheres of influence. As of the passage of the act, the process of licensing was to be managed not by the colleges, but by a new institution – the General Medical Council (GMC), whereas the colleges, despite the fact that they retained their independence, became only education and examination institutions. Therefore, it was possible to pass the examination in Glasgow, receive the GMC licence and practice in London or anywhere else in Great Britain.

Pursuant to the aforementioned act, the membership examination was also changed. The titles of "members" and "fellows" were introduced. The entrance examination to the college (RCP) was an examination for membership (thence the acronym: MRCP), whereas one could become a fellow in recognition of merit to the medical community. Fellows were elected among members through voting. The new examination was conducted for the first time in 1859. In 1867, the licentiate title was restored, which was reserved for medical students (Cooke, 1972).

The examination was formalised and, in this form, became the basis for the present-day MRCP(UK). At that time, it consisted of four parts: written essays from theoretical knowledge of physiology, pathology (each of them with translations from Latin and Greek) and an oral examination on the use and practice of medicine (RCP and Cuthbertson, 2008). The examinations took place in stages every three months. Exemption from the written examination was possible only for those candidates who were deemed particularly talented by the

college, yet it was not specified exactly what this term meant. In 1867, a change was introduced, resulting from a recommendation of the GMC to have the examinations assessed by a pair of examiners, and not one examiner as before (Cooke, 1972).

The publications of the “British Medical Journal” from 1879 and 1880 (GMC, 1879, 1880) show the requirements for candidates to the college, yet they do not provide information about the assessment procedures during the examination. It is believed that quality assurance during the examination was the examiners’ task. Candidates for college members had to be older than 25, have studied for at least five years, out of which four years at a renowned university acknowledged by the college. It is also known that in this period, the examination consisted of three parts and questions from new areas of medicine appeared.

Part I included written questions from anatomy and physiology and an oral examination on pathology and making medical preparations. This part lasted for a total of two days. Part II lasted for four days, in the course of which a written and oral examination was held on medicine, chemistry in pathology, pharmacy and toxicology; a written examination on women’s diseases and obstetrics, a written examination on anatomy and surgery, a practical examination at a hospital ward and an oral examination on surgery and obstetrics. A third examination lasted for five days in total and consisted of another series of written questions and a practical examination held at a hospital ward.

In 1893, the first note was published pertaining to the assessment procedure of candidates during the examination. Doctor William Dickinson, a retired censor (i.e. examiner) criticised the varied standards of assessing candidates and overly lenient marks, and suggested a unified method of assessing the candidates’ answers (Cooke, 1972). As a result of such criticism, in June 1894, the college introduced a new, point-based system of marks: a maximum of 50 points for every essay, 20 points for familiarity with language and 120 points for the oral examinations.

At that time, the colleges in Edinburgh and Glasgow introduced their own examinations. The examination in the Edinburgh college was conducted for the first time in 1881 and consisted of a written task in the area of medical practice and selected medical specialty (Fleming et al., 1974) and a practical examination consisting of long cases and short cases, which referred to whether the patient’s ailment was chronic or acute. The total duration of the examination was three days.

The college in Glasgow had its own examination system since the very beginning, yet after passage of the act of 1858, it did not hold examinations. For 27 years, it accepted anyone who could present a title from a reputable British or foreign university on the condition that the candidate received two-thirds of the fellows’ votes for membership. After 1885, due to charges on failing to fulfil quality standards, a new examination was introduced in Glasgow, based on the experiences of the Edinburgh college. Discussions among the members of the London and Scottish colleges about the quality of examinations in Scotland were made public in the “British Medical Journal” and offer interesting reading even today.

4.3.3 Beginning of the 20th century

New technological advances in medicine and changes in the design of the examination before the outbreak of WWII introduced much more serious changes.

The beginning of the 20th century brought new technological advances and the rapid development of medical knowledge: X-rays were discovered, as were hormones, and the first immunisation was developed. Initially, this did not result in changes in the MRCP examination, even though discussions were held on introducing the new scopes of knowledge and changes in the examination regulations. Between 1916 and 1924, the language examination gradually changed and finally in 1925, a separate examination was established. An assessment of language skills was completely abandoned in the 1960s.

The experiences of WWII greatly affected the way examinations were held. In the inter-war period, the passing rate of the MRCP examination was low. In 1933, only 8 persons passed out of 75 candidates. In 1939, 25 out of 130 candidates passed the examination. According to the statistics, the passing rate in that period

amounted to approx. 25% (Cooke, 1972). Right before the outbreak of WWII, in 1937, a re-designed examination was introduced. At that time, the MRCP consisted of two written parts (four open-ended questions in each part) and two oral parts. No sample sheets were published, so it is hard to determine how these questions differed from the previous ones, yet the examination became shorter.

During WWII, in response to the growing demand for physicians and broader access to higher education, the number of persons willing to join the college greatly increased, which however, lacked the needed group of examiners. Their number was increased as late as in 1949.

4.3.4. The 1960s and the increase in recruitment to higher education

In the post-war period, a pre-selection test was introduced to the MRCP(UK), allowing the efficiency to be significantly increased of conducting the examinations and assessing candidates' knowledge before the key practical part. Shortly afterwards, one joint examination under the MRCP(UK) name was introduced, where long open-ended questions were replaced by short ones divided into groups. Familiarity with cases of common diseases was emphasised rather than rare and complex ones.

Until 1963, the examination consisted of several written and oral parts, as was established in 1858, and only the written parts became less demanding. At that time, it was commonly believed that the examination confirmed a person's familiarity with rare diseases rather than common ones. Another problem was the passing rate, which was lower than the demand of society for medical services at that time. The small number of candidates accepted for specialty training resulted directly from the structure of the examination. Everyone who fulfilled the formal criteria took all parts of the MRCP examination, which resulted in the identical number of candidates taking the written and practical tests, thereby reducing the efficiency of the entire system along with the efficiency of examining.

Due to the fact that the practical part of the examination was considered vitally important, candidates' knowledge had to be confirmed before this stage. a rational decision was dividing the MRCP into two parts, whereas only the second part consisted of a written test and oral examination. Part I of this examination became the pre-selection test. In the new structure, long open-ended questions were no longer used and replaced by shorter ones with two possible answers: "yes" or "no." New advancements in technology, i.e. the computer from Newcastle University, were used to assess this test. Examination questions were still prepared by college examiners. In 1966, a bank of questions was created and a special group was set up in the RCP to work on devising new questions. According to the available literature, psychometric knowledge was used for the first time as late as in 1967. At that time, the discriminatory power and the predictive value of the examination was evaluated.

In 1969, part II of the examination was also modified. Instead of long open-ended questions, which were considered the weakest and most time-consuming form of the examination, short open-ended questions were introduced, divided into three groups: history of cases, data interpretation, and graphic materials. The answers to these questions were personally assessed by the examiners (RCP and Cuthbertson, 2008), but the answer key was previously agreed to by the persons assessing the work, whereas the marks were determined on the basis of the probability of the occurrence of possible answers. An example of such a question is included in Fleming's publication (Fleming et al., 1974).

The oral examination, i.e. practical examination in part II, was still used, as it offered the examiners the possibility of directly assessing the candidate's knowledge and skills. However, its form changed. Traditionally, candidates had to address one long (chronic) medical case and several short (acute) ones. Due to the fact that there are few diseases that are characterised by a complex course, the long cases used in the examination were commonly known by the candidates. Because of this, it could not be considered reliable, as the medical diagnosis proposed by the candidates during the examination did not reliably present their actual medical skills. Therefore, long cases were completely eliminated from the examination questions. Furthermore, as mentioned before, the MRCP was considered an examination on rare diseases or atypical cases. With the new design, the examination covered cases of common diseases, but sometimes offering atypical symptoms. This change was meant to expand the pool of examination cases. In its new form, the

MRCP fulfilled the requirements of the three colleges of physicians, i.e. London, Glasgow and Edinburgh. In 1969, the MRCP became a joint examination for all of them and changed its name to the MRCP(UK).

Between 1950 and 1970, part III was added to the MRCP for some specialties, e.g. psychiatry or paediatrics, which were, however, eliminated at the moment when separate colleges for these specialties were established.

4.3.5. From the 1970s to the present

Technological changes in the 1970s not only facilitated confirmation of possessed competences, but also enabled the psychometric evaluation of the examination, which resulted in the long-term and detailed verification of the accuracy and reliability of the MRCP(UK). The beginning of the 21st century brought a considerable change through the introduction of the practical assessment of clinical skills (PACES), which was also used to confirm communication skills, increasing in importance in the medical profession.

The frequency of introducing changes to the structure of the examination and its scope in this period was significant. This reflected advancements in medical science, technology as well as the social and political transformations that were taking place. So far, examinations were evaluated personally by examiners. Despite defined criteria and attention to the consistency of marks, it was a time-consuming process, requiring the engagement of a number of people. Furthermore, the calculation of the statistics and parameters used to evaluate the quality of the MRCP(UK) required considerable effort. The availability of computers resulted not only in the more efficient calculation of examination results, but also greatly facilitated the performance of psychometric analyses. 1984-2015 was also a time of continually verifying the accuracy of the MRCP(UK) examination for the purpose of maintaining its reliability and quality.

4.3.5.1. Format of questions and marking

Results of the analyses of test examinations showed that questions having a “yes/no” format made it easier for candidates to accidentally provide the correct answers. As a result, at the beginning of the 1970s, the examination sheets were modified and negative marks were introduced to the marking system to discourage examination participants from guessing.

In 1984, it was determined that part I would be composed of 60 questions with five sub-points in the “yes/no” format, which in total provided 300 evaluated test items, and this format was used until 2002.

Part I was developed with the aim of pre-selecting candidates, therefore part II could refer to more detailed issues and became an examination focused on confirming practical clinical knowledge. The results of this part consisted of marks from the test and marks from the clinical examination. Test questions (analogously to part I) began referring to such aspects as: the evaluation of a patient’s situation, the ability to read test results, diagnosis, making decisions and prognoses for the patient. The clinical part did not change in comparison to the previous years.

4.3.5.2. Psychometric evaluation

An important element of quality assurance for the examination is psychometric evaluation. Between 1984 and 2001, the reliability of the examination measured with the coefficient of reliability (Cronbach's alpha) was approx. 0.87 (McManus *et al.*, 2003) and gradually increased to reach a level of 0.92. The reliability of part II was lower and between 2002 and 2005, its coefficients of reliability ranged from 0.73 to 0.83 (Tighe, McManus, Dewhurst, Chis and Muckow, 2010) of the Standard Error of Measurement (SEM). Until 2005, part II consisted of two tests with 100 questions each, yet due to its low reliability, the number of questions grew to 270, which is consistent with the Spearman-Brown formula (cf. Hornowska, 2005). Eventually, the reliability of the measurement for part II was deemed satisfactory and reached the level of 0.90.

As part of the evaluation of the quality of the examination, accuracy was also studied. Not only hypotheses relating to social issues were verified, for example, potential discrimination due to sex (Dewhurst *et al.*, 2007,

McManus and Elder et al., 2008), but also the correlation between the results of the MRCP(UK) examination and the effects of the work of doctors or the results of further training (the so-called predictive accuracy; Levy et al., 2011a; Ludka-Stempień, 2015). The purpose of such analyses was to ensure the full transparency of the institution and the MRCP(UK) examination, as well as its reliability.

Some of the performed analyses had considerable impact, if not on the structure of the examination, then on the examination procedure. For example, a project implemented by the McManus team (McManus et al., 2005) analysing the skills of graduates of individual years who took the examination between 1985 and 2001 showed that the university graduates of 1985-1996 had much higher competences than the graduates of subsequent years. The reasons for this were not unequivocally presented in the article because this could potentially have been affected by a number of factors. Nevertheless, the results justified the RCP's use of new methods to determine the threshold mark for passing the examination rather than by relying on norm-referencing.

It was initially determined that the minimum threshold marks for the tests (part I and II) would be a fixed percentage of correct answers. Yet this resulted in the candidates challenging the reliability of the examination, because the marking system was completely detached from both the level of difficulty of the examination as such, and the level of skills in the population which it was meant to examine. Given the fact that the questions were changed in subsequent test editions, this method for determining the threshold made forecasting the number of candidates who would pass it impossible. An alternative approach was not to determine the minimum mark to be received at the examination, but the maximum number of candidates who could pass the examination. The limit was set at 35% of all persons taking the examination (norm-referencing). However, this solution still did not guarantee that candidates accepted to the college would have a sufficient level of knowledge, and, as shown by the McManus study mentioned in the prior paragraph, the level of knowledge and skills of candidates completing medical studies was dropping.

In relation to the above, the Angoff method was introduced to determine the passing rate threshold, and subsequently, the Hofstee method was used. A detailed description of both methods can be found in Norcini (2003). The Angoff method relies on a group of competent examiners assessing the skills and knowledge that a candidate should have in order to receive the threshold mark for passing the examination (borderline candidate) and - taking the questions into account - the probability that such a candidate will answer every question. On this basis, the average of the marks of the examination committee members is calculated, which offers the borderline mark.

The Hofstee method does not assess all test items, but indicates the range of threshold marks allowing the examination to be passed and the percentage of the desired number of candidates who are going to have a positive result (Hofstee, Berge and Hendriks, 1998; Norcini, 2003). The Hofstee method functioned until the end of 2008 for part I and until the end of 2009 for part II.

Eventually, the best and the most-recent method for determining the threshold mark was introduced – statistical equating. This method is quite complex, yet it was described well by Skaggs and Lissitz (1986) and Moses et al. (2010) in *The Use of Two Anchors in Non-Equivalent Groups with Anchor Test NEAT*. It relies on the parameters of individual items and allows the difficulty of every question and the potential of the group of candidates to be taken into account.

Another psychometric study relating to multiple attempts at passing the examination has shown that there is a group of candidates for whom these multiple attempts may lead to passing it “by luck” and receipt of the qualification in the absence of actual competence (McManus and Ludka, 2013). The effect of this study was limiting the maximum number of attempts at passing the examination's individual parts to six in 2014.

4.3.5.3. Separation of PACES and assessing its accuracy and reliability

In 2001, part II was officially divided into a written test and PACES (Practical Assessment of Clinical Examination Skills), thus PACES became a separate part of MRCP(UK). Until recently, it was possible to take this part only after having passed part II.

PACES was introduced as the result of broader changes in the thinking about examining actual skills and a public discussion on this issue. Accusations were raised against the MRCP(UK) that the structure of the examination and the probability of passing it were similar to the rules of a game of chance (London, Parry, Clesham and O'Brien, 1994). The purpose of the new form of the clinical examination was setting it in the reality of modern medicine and the thinking based on scientific evidence, and adjusting it to modern methods of assessing skills that were already in use at universities (Dacre, Besser and White, 2003). Therefore, PACES was modelled on the objective, structured clinical examination process (OSCE), which in the 1990s became a popular model for assessing the skills of students at medical universities. OSCE is a series of brief tests and clinical tasks developed to more quickly and efficiently assess the skills of medical students (Harden, Stevenson, Downie and Wilson, 1975). As for the scope of the assessed clinical skills, little has changed with the exception of placing greater emphasis on communication and history taking skills, which was also reflected in the changes made to *Good Medical Practice*.

Since its onset, PACES has had a formula similar to the current one. In 2009, modifications were introduced, primarily consisting of a change from the hitherto assessment of specific skills to general competences, for example: the ability to communicate with the patient, addressing the patient's concerns, performance of a physical examination. The candidate must receive a minimum number of points on every scale determining competence and not, as before, simply receive a good score in three out of five tasks, which allowed candidates to compensate for deficiencies in communication by performing a good physical examination or vice versa (Elder et al. 2011).

Due to the practical nature of the examination, evaluating its quality was a much more difficult task than in the case of tests. However, the reliability of its results was also confirmed. The test reliability coefficient calculated by McManus (McManus, Thompson and Mollon, 2006) was 0.82, which was a satisfactory, yet not an excellent result. The results presented by Dacre et al. (2003) indicated the high reliability of the marks given by examiners, which was evidence of a high level of standardisation of the examination procedure. In reference to the results of these studies, the awarded mark differed by more than two points only in the case of 2.2% of the examiners working in pairs, and 60.7% of the examiners were completely unanimous in their marks (Dacre et al., 2003). Furthermore, the 2006 McManus study showed that 87% of the variance in examination results was caused by actual differences in the competence level of the candidates, 1% of the differences in the assessments could be ascribed to the specific nature of the performed task, whereas 12% resulted from the individual way in which each examiner assessed the candidates. As McManus et al. reported (2006): "(...) some examiners being more stringent and requiring a higher performance than other examiners who are more lenient. Although the problem has been known qualitatively for at least a century, we know of no previous statistical estimation of the size of the effect in a large-scale, high-stakes examination. Here we use FACETS to carry out a multi-facet Rasch modelling of the paired judgements made by examiners in the clinical examination (PACES)". Such an undesired effect of the differences among examiners was analysed in terms of the impact on the overall results of the examination for individual candidates. It turned out that the natural predilection for assessing more stringently or more leniently impacted the result of the examination (it could have been potentially different) for only 4.1% of the candidates. 2.6% of the candidates should have passed the examination, whereas 1.5% received unjustly higher results, which allowed them to pass.

In the case of the oral examination, important factors which may affect the results are the sex and ethnic origin of the candidates and examiners and the risk of covert discrimination resulting from this. In 2007, studies were conducted (Dewhurst et al., 2007) which showed that there is a slight effect of the interaction between the ethnic origin of the candidates and the examiners. It was observed that ethnic minority examiners assessed minority candidates more leniently, but this was not observed for non-minority examiners in reference to any group represented by the candidates. This effect only occurred in tasks where communication was assessed, whereas the existence of such an effect was not exhibited in tasks relating to clinical skills, which may indicate certain cultural factors that only emerge in situations of direct contact with persons from similar cultural circles.

According to further studies on the impact of ethnic origin and sex on examination results (Dewhurst et al.,

2007), PACES was the hardest part for men of non-European ethnic origin and they performed much worse than the group of ethnically European men and ethnically non-European women. Women performed better than men, whereas this effect was ascribed not to procedural factors, but to the actual skills of the candidates. Similar results were also achieved by Bessant et al. (2006).

Due to complaints lodged by general physicians in the college (MRCGP) and the court petitions of ethnic minorities following them in 2012 (The British Association of Physicians of Indian Origin, BAPIO), the potential impact of ethnic origin on the results of MRCP(UK) was reviewed once again. The studies of 2013 verified the impact of ethnic origin and sex on examination results, but showed that these factors did not affect the marks received. Therefore, no further actions were taken relating to the structure of the examination that would take this aspect into account (McManus, Elder and Dacre, 2013). Other studies (Wakeford, Denney, Ludka-Stempień, Dacre and McManus, 2015) also confirming the accuracy of the MRCP(UK) examination also showed that potential differences in the results of candidates from various ethnic groups are rather the effect of differences in skills and knowledge than the result of discrimination against any of the groups. Earlier studies indicated experiences, cultural factors and language skills as the causes for such differences (Esmail and Roberts, 2013; Slowther, Lewando Hundt, Taylor and Purkis, 2009).

The hypothesis about cultural factors was also verified in the study conducted by Verma's team (Verma, Griffin, Dacre and Elder, 2016). They found that the failure of persons who were educated outside the United Kingdom, most of whom came from other than European ethnic groups, was affected by poor results in five examined aspects: ability in understanding instructions, ability to cope with the patient's fears, building relationships, structuring an appropriate response, explaining a situation and planning.

Researchers indicate that some cultural norms may negatively impact the direct physician-patient relationship. For example, in cultures where there is a significant distance in the relationship between a person with authority and a person without authority, physicians are less apt to properly provide patients with information.

5. Comments

The MRCP(UK) examination must be carefully evaluated with respect to reliability, not only due to the fact that it is a compulsory stage in an individual's career path, but also because the profession of physician is one of public trust, and a lack of competence could lead to risks for patients. Political and cultural determinants resulted in the fact that the medical industry has historically secured a certain independence in the area of assessing the competences of physicians providing services, conducting training and granting professional licences. The right to such autonomy was acquired in exchange for ensuring the high quality of provided services. However, the social trust which the medical industry had enjoyed in the last 20 years or so has decreased and the quality of medical services has been criticized. Maintaining the elementary independence of colleges requires that both the process of training doctors, as well as the process of confirming their competences through examinations is transparent and fulfils specific standards.

Changes introduced to the procedure of conducting examinations or the structure of the MRCP(UK) served the aims listed above. Most of them have been introduced since the 1960s in response to the far-reaching changes in the thinking about examinations, which resulted both from technological advances, the availability of the calculating capacity of computers, and progress in the area of psychometrics. Yet first and foremost, these changes resulted from the perception of the role and responsibilities of doctors in society and the growing understanding of patients' rights.

The purpose of maintaining high requirements of MRCP(UK) candidates is to ensure the best possible medical care to patients and to sustain a high level of social trust in the medical sector. The result of research projects on the MRCP(UK) confirmed, among others, the high compliance of examiners' scoring of the PACES examination, the lack of discrimination due to ethnic or national origin, confirmation that all parts of the MRCP(UK) are in line with the results of other medical examinations and workplace assessments, and that poor results in the MRCP(UK) foreshadow later potential problems in performing the duties of the medical

profession. In its present form, the examination uses various methods to assess relevant skills and at the same time, it is standardised, reliable and accurate and fulfils the international standards for educational and psychological tests. The great emphasis placed in the Integrated Qualifications System on the reliability and accuracy of the assessment of the learning outcomes required for a qualification is therefore justified and desirable.

On a higher level, compliance with the standards means that the MRCP(UK) fulfils its purpose, which is admission to specialty training of those physicians whose knowledge, skills and professional approach provide a good prognosis for their subsequent performance in the medical profession, in line with the professional standards set forth in *Good Medical Practice*. Despite the fact that the perceived aspects characterising a good doctor have changed greatly in the last several decades - dictated both by the progressive professionalization of services, as well as the media reports of incompetent doctors - the standards being set for new generations of physicians have become objectively higher. Apart from expanding professional ethics, guidelines pertaining to relations with colleagues, and increasing emphasis on the continuous development of competences, special importance has been placed on the ability to communicate with patients and other persons involved in their care, which resulted directly from the needs and complaints lodged by patients. A change in the thinking about these issues led to the inclusion of the above-listed skills in the medical curriculum, and in the case of the MRCP(UK), resulted in the changes introduced, for example, to PACES.

Ensuring the quality of the examination imposes on the college – as the examining institution – not only the task of monitoring the entire process of educating doctors, but also broad cooperation with other entities, including the ones that are directly responsible for training and the ones that supervise the entire education process. These two functions, as has been shown, are separate in Great Britain. This allows for greater transparency of the process of education and examination, as well as for ensuring the quality of such services and thus increases society's trust of the medical sector.

BIBLIOGRAPHY:

- Bessant, R., Bessant, D., Chesser, A., Coakley, G. (2006). Analysis of predictors of success in the MRCP (UK) PACES examination in candidates attending a revision course. *Postgraduate Medical Journal*, 82(964), 145–149.
- Clark, G. (1965). History of the Royal College of Physicians of London. *British Medical Journal*, 1(5427), 79.
- Cooke, A. M. (1972). *a History of the Royal College of Physicians of London*, tom 3. Oxford: Clarendon Press.
- Dacre, J., Besser, M., White, P. (2003). MRCP(UK) PART 2 Clinical Examination (PACES): a review of the first four examination sessions (June 2001 - July 2002). *Clinical Medicine*, 3(5), 452–459.
- Dewhurst, N. G., McManus, I. C., Mollon, J., Dacre, J. E., Vale, A. J. (2007). Performance in the MRCP(UK) Examination 2003-4: analysis of pass rates of UK graduates in relation to self-declared ethnicity and gender. *BMC Medicine*, 5, 8.
- Elder, A., McAlpine, L., Bateman, N., Dacre, J., Kopelman, P., i McManus, C. (2011). Changing PACES: Developments to the examination in 2009. *Clinical Medicine, Journal of the Royal College of Physicians of London*, 11(3), 231–234.
- Esmail, A., i Roberts, C. (2013). Academic performance of ethnic minority candidates and discrimination in the MRCGP examinations between 2010 and 2012: analysis of data. *BMJ: British Medical Journal*, 5662(September), 1–10.
- Fleming, A. P. R., Manderson, W. G., Matthews, M. B., Sanderson, P. H., Stokes, J. F., Fleming, P. R. (1974). Evaluation of an Examination: M.R.C.P. (U.K.). *Medical Education*, 2(5910), 99–107.
- GMC. (1879). Regulations of the general medical council and medical licensing bodies (Session 1879-80). *British Medical Journal*, 397–438.
- GMC. (1880). Regulations of the general medical council and medical licensing bodies (Session 1880-81). *British Journal of Anaesthesia*, (September), 415–454.

- GMC. (2013). *Good Medical Practice: Working with doctors working for patients*. London: GMC.
- Harden, R. M., Stevenson, M., Downie, W. W., i Wilson, G. M. (1975). Assessment of clinical competence using objective structured examination. *British Medical Journal*, 1(5955), 447–451.
- Hofstee, W. K. B., Berge, J., i Hendriks, A. (1998). How to score questionnaires. *Personality and Individual Differences*, 25(5), 897–909.
- Hornowska, E. (2005). *Testy psychologiczne. Teoria i praktyka*. Warszawa: Wydawnictwo Naukowe Scholar.
- Levy, J. B., Mohanaruban, A., Smith D. (2011). The relationship between performance in work place based assessments, MRCP exam and outcome from core medical training. *Medical Education*, 45. Dodatek 5.
- London, D., Parry, R., Clesham, G., O'Brien, E. (1994). Game, set, and MRCP. *The Lancet*, 344(October), 1159.
- Ludka-Stempień, K. (2015). Predictive validity of the examination for the Membership of the Royal Colleges of Physicians of the United Kingdom. University College London. Pobrano z <http://discovery.ucl.ac.uk/1467148/3/PhD Thesis v15 final for submission.pdf>
- McManus, I. C., Elder, A. T., i Dacre, J. (2013). Investigating possible ethnicity and sex bias in clinical examiners: an analysis of data from the MRCP(UK) PACES and nPACES examinations. *BMC Medical Education*, 13(1), 103.
- McManus, I. C., Mooney-Somers, J., Dacre, J. E., i Vale, A. J. (2003). Reliability of the MRCP(UK) Part i Examination, 1984-2001. *Medical Education*, 37(7), 609–611.
- McManus, I. C., Thompson, M., i Mollon, J. (2006). Assessment of examiner leniency and stringency ('hawk-dove effect') in the MRCP(UK) clinical examination (PACES) using multi-facet Rasch modelling. *BMC Medical Education*, 6, 42.
- Moses, T., Deng, W., Zhang, Y. (2010). The Use of Two Anchors in Nonequivalent Groups With Anchor Test (NEAT) Equating. Educational Testing Service, Princeton.
- Norcini, J. J. (2003). Setting standards on educational tests. *Medical Education*, 37(5), 464–469.
- RCP. (2016b). MRCP(UK) PACES Examination. Trainee Examiner's Guide. RCP. Pobrano z <https://www.mrcpuk.org/sites/default/files/documents/trainee-examiner-guide.pdf>
- RCP. (2016c). MRCP (UK) Part 2 Clinical Examination (PACES) Examiners' Guide Omnibus. RCP. Pobrano z <https://www.mrcpuk.org/sites/default/files/documents/Examiners Omnibus 2016 v1.1.pdf>
- RCP. (2016d). Regulations and Information for MRCP (UK) Candidates - 2017 Edition. London: RCP. Pobrano z <https://www.mrcpuk.org/sites/default/files/documents/MRCP%28UK%29-Regulations-2016-17.pdf>
- Skaggs, G., i Lissitz, R. W. (1986). IRT Test Equating: Relevant Issues and a Review of Recent Research. *Review of Educational Research*, 56(4), 495–529.
- Slowther, A., Lewando Hundt, G., Taylor, R., Purkis, J. (2009). Non UK qualified doctors and Good Medical Practice: the experience of working within a different professional framework, (February), 1–81. Pobrano z <http://wrap.warwick.ac.uk/48945/>
- Tighe, J., McManus, I. C., Dewhurst, N. G., Chis, L., Mucklow, J. (2010). The standard error of measurement is a more appropriate measure of quality for postgraduate medical assessments than is reliability: an analysis of MRCP(UK) examinations. *BMC Medical Education*, 10, 40.
- Verma, A., Griffin, A., Dacre, J., i Elder, A. (2016). Exploring cultural and linguistic influences on clinical communication skills: a qualitative study of International Medical Graduates. *BMC Medical Education*, 16(1), 162.
- Waddington, I. (1973). The struggle to reform the Royal College of Physicians, 1767-1771: a sociological analysis. *Medical History*, 17(2), 107–26.
- Wakeford, R., Denney, M., Ludka-Stempień, K., Dacre, J., i McManus, I. C. (2015). Cross-comparison of MRCP & MRCP(UK) in a database linkage study of 2,284 candidates taking both examinations: assessment of validity and differential performance by ethnicity. *BMC Medical Education*, 15(1), 1.

