Reumatología Clínica

Original Article

The Method Used to Set the Pass Mark in an Objective Structured Clinical Examination Defines the Performance of Candidates for Certification as Rheumatologists☆

Virginia Pascual-Ramos,* Ana Guilaisne Bernard-Medina, Diana Elsa Flores-Alvarado, Margarita Portela-Hernández, María del Rocío Maldonado-Velázquez, Luis Javier Jara-Quezada, Luis Manuel Amezcua-Guerra, Nadina E. Rubio-Judith López-Zepeda, Everardo Álvarez-Hernandez, Miguel Ángel Saavedra, César Alejandro Arce-Salinas, Consejo Mexicano de Reumatología (CMR)

Consejo Mexicano de Reumatología, Mexico City, Mexico

A B S T R A C T

Background: The Mexican Accreditation Council for Rheumatology certifies trainees (TR) on an annual basis using both a multiple-choice question (MCQ) test and an objective structured clinical examination (OSCE). For 2013 and 2014, the OSCE pass mark (PM) was set by criterion referencing as ≥ 6 (CPM), whereas overall rating of borderline performance method (BPM) was added for 2015 and 2016 accreditations. We compared OSCE TR performance according to CPM and BPM, and examined whether correlations between MCQ and OSCE were affected by PM.

Methods: Forty-three (2015) and 37 (2016) candidates underwent both tests. Altogether, OSCE were integrated by 15 validated stations; one evaluator per station scored TR performance according to a station-tailored check-list and a Likert scale (fail, borderline, above range) of overall performance. A composite OSCE score was derived for each candidate. Appropriate statistics were used.

Results: Mean ($\pm$standard derivation [SD]) MCQ test scores were 6.6 ± 0.6 (2015) and 6.4 ± 0.6 (2016) with 5 candidates receiving a failing score each year. Mean ($\pm$SD) OSCE scores were 7.4 ± 0.6 (2015) and 7.3 ± 0.6 (2016); no candidate received a failing CPM score in either 2015 or 2016 OSCE, although 21 (49%) and 19 (51%) TR, respectively, received a failing BPM score (calculated as 7.3 and 7.4, respectively). Stations for BPM ranged from 4.5 to 9.5; overall, candidates showed better performance in CPM.

In all, MCQ correlated with composite OSCE, $r = 0.67$ (2015) and $r = 0.53$ (2016); $P \leq .001$. Trainees with a passing BPM score in OSCE had higher MCQ scores than those with a failing score.

Conclusions: Overall, OSCE-PM selection impacted candidates' performance but had a limited affect on correlation between clinical and practical examinations.

© 2016 Sociedad Española de Reumatología y Colegio Mexicano de Reumatología. Published by Elsevier España, S.L.U. All rights reserved.

El método para establecer el punto de corte en el examen clínico objetivo estructurado define el desempeño de los candidatos a la certificación como reumatólogo

R E S U M E N

Antecedentes: El Consejo Mexicano de Reumatología certifica candidatos mediante una evaluación teórica (ET) y un examen clínico objetivo estructurado (ECOE). En 2013 y 2014, el punto de corte para acreditar el ECOE se estableció por criterio (PC ≥ 6); a partir del 2015, también se estableció por el método del desempeño limítrofe (PDL). Se comparó el desempeño de los candidatos con ambos puntos de corte y examinó su impacto en la correlación entre la ET y el ECOE.

Material y métodos: En 2015 y 2016, respectivamente, 43 y 37 candidatos aplicaron; ambos ECOE se integraron con 15 estaciones; un evaluador por estación calificó la lista de cotejo y el desempeño global del candidato mediante una escala de Likert (inferior, limítrofe y superior). A cada candidato se le asignó una calificación global del ECOE.

Resultados: El promedio (±DE) de la ET fue de 6,6 (±0,6) en el 2015 y de 6,4 (±0,6) en el 2016; 5 candidatos/año no acreditaron. El promedio (±DE) del ECOE fue 7,4 (±0,6) y 7,3 (±0,6), respectivamente; todos acreditaron de acuerdo con el PC, mientras que 21 (49%) y 19 (51%) no lo hicieron de acuerdo con el PDL (7,3 en 2015 y 7,4 en 2016). Los PDL para cada estación variaron.

La ET correlacionó con el ECOE. Los candidatos con ECOE acreditado (por PDL) calificaron mejor en la ET que su contraparte.

Conclusiones: El método para establecer el punto de corte del ECOE afecta al desempeño de los candidatos a una certificación, pero no impacta a la correlación entre la ET y el ECOE.

Introduction

Councils that certify health specialists are committed to society, to ensure that certified physicians possess the clinical skills and necessary knowledge to practice their profession. As a part of this commitment, they develop and apply evaluation tools, establish pass marks and, finally, draft critical decisions, not only for those involved but also for the medical community and society. Thus, the choice of the pass mark must be robust from the methodological point of view, since this varies substantially depending on the method applied to calculate it.[1]

For years, the Mexican Accreditation Council for Rheumatology (CMR) has certified all the residents who have completed their studies as rheumatologists on an annual basis; certification is carried out at the end of training and consists of a written test with multiple-choice questions that evaluates theoretical knowledge (MCQ) and an examination to test clinical skills. The latter was traditionally based on a single case; due to the limitations inherent to said assessment tool modality,[2,3] in 2013 we implemented and subsequently began to apply an objective structured clinical examination (OSCE) as part of the certification process.[4]

For certification in 2013 and 2014, the OSCE pass mark for accreditation was established "by criterion" as being 6 or more (in a scale of 0 to 10, in which 0 represented the worst performance); a session held expressly for this purpose was devoted to discuss and agree on the minimum number of items on a validated checklist to thus consider whether the skills of a rheumatologist were adequate and safe. Beginning in 2015, the evaluators assigned to each OSCE station were instructed and trained to complete the checklist, as well as to evaluate the general performance of the candidate using a Likert scale for the purpose of establishing the OSCE pass mark on the basis of the overall borderline performance (BPM).[5]

The objectives of the exercise were:

1. To compare the OSCE performance of the candidates for certification as rheumatologists, using pass marks established by 2 different methods, by criterion (CPM) and by the BPM method.
2. To examine whether the correlation between the MCQ and the OSCE is affected by the pass mark selected.

Material and Methods

In Mexico, there are 16 centers accredited to prepare specialists in rheumatology. In 2015 and 2016, 43 and 37 candidates, respectively, applied for certification in rheumatology over 2 consecutive days; all of the candidates had completed a training program with a duration of at least 4 years in their respective educational institutions and had been recommended by their professors.

The 2 versions of the MCQ consisted, respectively, of 222 and 200 questions, mostly presented in the format of case reports and posed by experienced certified rheumatologists, and reviewed by a council subcommittee formed by 4 certified rheumatologists. Both OSCE circuits were comprised of 15 stations designed by members of the council; each station was validated by a subcommittee of at least 6 certified rheumatologists who had not been involved in the design of the stations, provided that at least 80% of the evaluators approved the inclusion of each item on the checklist; each station had a duration of 8 min and the circuits included 4 rest stations. An external evaluator was assigned to each station. He or she was duly trained to score the checklist and the Likert scale concerning the overall performance of the candidate (fail, borderline and above range). The checklists of each station included a number of items that ranged from 5 to 21.

The following pass marks were established for those 2 years: the process of certification in 2015 required a score ≥ 5.7 for the MCQ and ≥ 6 for the OSCE; the process of certification in 2016 required a score ≥ 6 for both the MCQ and OSCE evaluations. In all cases, the possible maximum score was 10.

The BPM was calculated as follows: for each station, we selected the checklists of the candidates scored as "borderline" according to the Likert scale for overall performance. Then, we determined the mean score obtained in those checklists, which was the BPM for the station being evaluated. Finally, we obtained the BPM for the overall OSCE by averaging the BPM of the 15 stations.

Each candidate was given a score for BPM and another for the OSCE (overall), by averaging the scores for the 15 stations.

Statistical Analysis

Descriptive statistics were employed and the tests appropriate for the variable distribution were utilized. For construct validity, we correlated the BPM and OSCE scores using the Pearson correlation coefficient; likewise, we compared the BPM scores of those whose OSCE scores were above and below the pass mark with the Mann–Whitney *U* test. The analyses were carried out with the SPSS/PC v20 statistical package.

Results

Table 1 summarizes the relevant data of the certification processes of 2015 and 2016. The average BPM scores were similar in the 2 years and close to 6.5, as were the percentages of candidates whose scores were below the pass mark, 12% and 14%, respectively.

The scores of the candidates in the OSCE were higher than those of the BPM in both years and close to 7.4; no candidate failed the OSCE according to the CPM. In both scores, the BPM of the OSCE was calculated to be 7.3 (in 2015) and 7.4 (in 2016), respectively.

Table 1
Results of the Certification Processes in 2015 and 2016.

	2015 No. = 43	2016 No. = 37
MCQ (mean ± SD)	6.6 ± 0.6	6.4 ± 0.6
Candidates with a failing score in MCQ, n (%)	5 (12)	5 (14)
Overall OSCE (mean ± SD)	7.4 ± 0.6	7.3 ± 0.6
Candidates with a score of OSCE < CPM, n (%)	0	0
Overall BPM	7.3	7.4
Candidates with a score of OSCE < BPM, n (%)	21 (49)	19 (51)
MCQ score in candidates with OSCE ≥ BPM, median (Q25–Q75)	7.1 (6.4–7.3)	7.8 (7.5–8)
MCQ score in candidates with OSCE < BPM, median (Q25–Q75)	6.3 (5.9–6.7)	6.9 (6.5–7.2)

BPM, borderline performance method; CPM, pass mark by criterion; MCQ, written test with multiple-choice questions that evaluates theoretical knowledge; OSCE, objective structured clinical examination; Q, quartile; SD, standard deviation.

For this, we used the checklists of 30 and 26 candidates, respectively, scored as being borderline. In accordance with the BPM, the percentage of candidates who failed increased to about 50% in both years (Table 1).

Table 2 shows the number of items on the checklist for each station, the scores of the candidates in each station, the BPM of each station and candidate performance in accordance with the CPM and the BPM for certification in 2015. Similar data were encountered in the certification process of 2016 (not shown). In general, there is a substantial variation in the BPM in each station, from 4.5 for station number 2 to a value of 9.5 for number 10. The performance of the candidates in each station was better when the CPM was used rather than the BPM method, except in the case of station number 2.

Figs. 1 and 2 show that the correlation between MCQ and OSCE was significant and moderate in both certification processes ($r = 0.67$ and $r = 0.53$, for 2015 and 2016, respectively; $P = .001$ in both cases). Similarly, the candidates with OSCE scores ≥ BPM had a better score in the MCQ than those with a score in the OSCE < BPM, as is shown in Table 1.
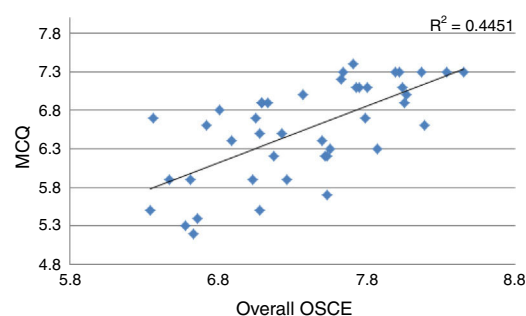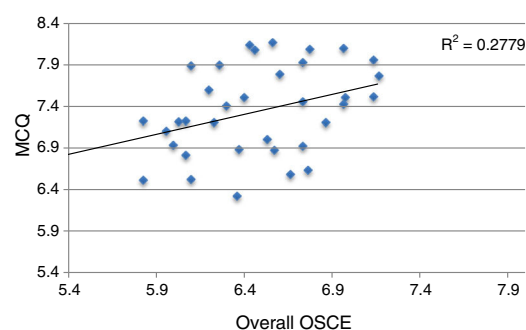
Discussion

Society at the present time has a growing interest in the performance of physicians and demands that health workers be adequately prepared to exercise their profession. This process

Table 2
Performance of the Candidates per Station in Accordance With the Pass Mark by Criterion and Borderline Performance Method During the Certification Process of 2015.

Station	Score: mean ± SD	BPM	No. (%) candidates with a score ≥ CPM	No. (%) candidates with a score ≥ BPM
1 (5)	7.6 ± 1.1	7.7	40 (93)	21 (49)
2 (5)	4.8 ± 1.6	4.5	11 (26)	18 (42)
3 (6)	6.8 ± 1.8	6.7	35 (81)	27 (63)
4 (14)	6.1 ± 1	6.0	21 (49)	13 (30)
5 (14)	7.4 ± 1.3	6.9	37 (86)	28 (65)
6 (21)	7.7 ± 1.3	7.6	35 (81)	26 (61)
7 (9)	7.7 ± 2.3	8.6	26 (61)	23 (53)
8 (15)	6 ± 1.7	6.8	19 (44)	13 (30)
9 (14)	7.7 ± 1.8	7.5	31 (72)	26 (61)
10 (19)	9.7 ± 0.6	9.5	43 (100)	31 (72)
11 (10)	7.7 ± 1.5	7.7	36 (84)	23 (53)
12 (8)	9.0 ± 1.3	8.6	43 (100)	22 (51)
13 (8)	7.3 ± 1.8	7.2	36 (84)	22 (51)
14 (6)	8.7 ± 1.4	8.2	42 (98)	35 (81)
15 (17)	6.6 ± 0.9	6.5	30 (70)	19 (44)

BPM, borderline performance method; CPM, pass mark by criterion, SD, standard deviation.

Fig. 1. Correlation between the written test with multiple-choice questions that evaluates theoretical knowledge and the objective structured clinical examination (2015 certification). MCQ, multiple-choice questions; OSCE, objective structured clinical examination.

Fig. 2. Correlation between the written test with multiple-choice questions that evaluates theoretical knowledge and the objective structured clinical examination (2016 certification). MCQ, multiple-choice questions; OSCE, objective structured clinical examination.

when defined in terms of regulation is known as certification of medical competence. It implies assessment, and that the evaluation be summative. It also requires the establishment of scores above which the candidate will (or will not) achieve certification. There are several methods to define those pass marks; however, the availability of human and economic resources is a key element in the selection of some in particular. At the present time, there is no gold standard to evaluate medical competence, although OSCE has gained followers in recent years.

This report demonstrates that candidates for certification in rheumatology showed an extremely variable OSCE performance, depending on how the *a priori* establishment of the pass mark was carried out to define accreditation in the evaluation of clinical competence. This became evident when overall OSCE was considered for each candidate, as has been proposed by some authors, rather than to analyze the performance in each station.[5] In addition to the CPM method, we calculated the findings with BPM, for which each evaluator assigned to a station, directly and repeatedly, determined the general performance of the candidates using a Likert scale; this strategy has been proposed as an indispensable element in the attempt to not affect the reliability of the evaluation method.[6] The additional advantage of using scales to evaluate the general performance of the candidate (in contrast to checklists) is that they can include qualitative aspects such as efficiency and ease in the activities performed by the candidates, important aspects in their own right and that are difficult to examine by means of checklists.[7] In fact, there is evidence that overall evaluations are more reliable than checklists,[6,7] although it has been pointed out that it is convenient to integrate several evaluation methods in the case of clinical competence with important consequences.[8–10] The

differences highlighted by each evaluation tool explain the variable performance of our candidates with both pass marks.

The selected pass mark was not affected by the correlation between the MCQ and the OSCE. Our report confirms earlier studies that showed a moderate-to-good correlation between the 2 evaluations.[11,12] Some authors have proposed that the proper correlation between the MCQ and the OSCE supports the construct validity of the OSCE; combined with the latter, we observe that the candidates who were not accredited by the OSCE in accordance with the BPM had a poorer performance in the MCQ. Finally, the construct validity of the BPM can be inferred from the fact that candidate performance scored as borderline varied from one station to another, as shown in Table 2. We were concerned that the evaluators might use the checklist score to evaluate the overall performance of the candidates; had they done so as a consistent policy by the different evaluators, the BPM of the different stations would have been close to 5. It is worth mentioning that the correlation between the MCQ and the OSCE does not necessarily imply that the 2 tools evaluate the same knowledge and thus are redundant; although a substantial part of what the OSCE evaluates corresponds to the "knows" level of Miller's pyramid,[13] we consider that the use of both tools increases the information relative to candidate competence. Taking into account that this increase is limited and that the OSCE consumes considerable human and economic resources, it is necessary to optimize it mainly to evaluate the "shows how" level.

This report has limitations that we point out here. The implementation and application of an OSCE is complex and requires extensive and consolidated experience, which the CMR is in the process of acquiring. A single evaluator per station (rather than 2) scored the checklist and the candidates' overall performance. Although we observed an adequate correlation between the MCQ and the OSCE, we did not follow the recommendations of Matsell et al.,[14] who suggested classifying the OSCE stations in domains (knowledge, problem-solving, clinical skills and patient management) and correlating them with their equivalents in the MCQ. We could not correlate the OSCE score with the evaluations obtained during resident training because we had no access to the latter. The CMR developed an OSCE comprised of a limited number of stations, the duration of which was short (8 min), all of which affects the evaluation of those problems that require the integration of different skills, the fundamental objective in the process of certification. Finally, in each certification process, we had to apply 2 consecutive OSCE due to the limited number of offices, which may have affected the reproducibility of the results[15]; however, the physical space, the patients and the evaluators were the same in both circuits.

In our specialty, there are few studies that analyze the performance of an OSCE to evaluate clinical competence.[16–20] Those that do so in the academic and professional setting of a certification process are even more limited.[5,17] The most recent trend in (medical) education is based on competence. Thus, the OSCE has the potential to be a suitable tool to evaluate clinical skills.[21] However, in those processes in which the outcome has an important impact on the candidates and society, we recommend the use of additional, complementary tools. This requires a careful analysis of the results, which depend on the method employed to define the pass mark, as can be seen in the present report.

Finally, reflection on the tools used to evaluate clinical skills (such as the OSCE) should extend to the teaching of these skills in our specialty; priority is currently given to the traditional model of medical education, which has not been standardized throughout the different teaching hospitals in Mexico. It is characterized by the diversity of the populations attended, the health services to which the centers belong and their physical and human resources. This training model signifies a delay in the incorporation of innovative education technologies, such as simulation-based medical education, which has been found to favor the acquisition of a wide range of clinical skill as compared to the traditional model.[22]

In conclusion, the CMR has invested time, commitment and resources to confer transparency, professionalism and reliability to the process of certification in rheumatology; thus, it has developed robust tools for evaluation and has questioned its performance. We consider that this is the path to follow, despite the doubts and frustration that emerge when we report failures, and that the same reflection and effort should be extended to the model for training our specialists.

Ethical Disclosures

Protection of human and animal subjects. The authors declare that no experiments were performed on humans or animals for this study.

Confidentiality of data. The authors declare that no patient data appear in this article.

Right to privacy and informed consent. The authors declare that no patient data appear in this article.

Conflict of Interest

The authors declare they have no conflicts of interest.

Acknowledgments

References

1. Downing SM, Haladyna TM. Validity threats: overcoming interface with proposed interpretations of assessment data. Med Educ. 2004;38:327–33.
2. Wass V, van der Vleuten C. The long case. Med Educ. 2004;38:1176–80.
3. Nweble DI. The observed long-case in clinical assessment. Med Educ. 1994;25:369–73.
4. Pascual Ramos V, Medrano Ramírez G, Solís Vallejo E, Bernard Medina AG, Flores Alvarado DE, Portela Hernández M, et al. Performance of an objective structured clinical examination in a national certification process of trainees in rheumatology. Reumatol Clin. 2015;11:215–20.
5. Wilkinson TJ, Newble DI, Frampton CM. Standard setting in an objective structured clinical examination: use of global ratings of borderline performance to determine the passing score. Med Educ. 2001;35:1043–9.
6. Regehr G, Mac Rae H, Reznik RK, Szalay D. Comparing the psychometric properties of checklists and global ratings scales for assessing performance on an OSCE-format examination. Acad Med. 1998;73:993–7.
7. Ilgen JS, Ma IWY, Hatala R, Cook DA. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. Med Educ. 2015;49:161–73.
8. Schuwirth LW, van der Vleuten CP. The use of clinical simulations in assessment. Med Educ. 2003;37:65–71.
9. Wass V, van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet. 2001;357:945–9.
10. Newble D. Techniques for measuring clinical competence: objective structured clinical examination. Med Educ. 2004;38:199–203.
11. Van der Vleuten CP, van Luyk SJ, Beckers HJ. A written test as an alternative to performance testing. Med Educ. 1989;23:97–107.
12. Kissin EY, Grayson PC, Cannella AC, Demarco PJ, Evangelisto A, Goyal J, et al. Musculoskeletal ultrasound objective structured clinical examination: an assessment of the test. Arthritis Care Res. 2014;66:2–6.
13. Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990;65 Suppl. 9:S63–7.
14. Matsell DG, Wolfish NM, Hsu E. Reliability and validity of the objective structured clinical examination in paediatrics. Med Educ. 1991;25:293–9.
15. Van der Vleuten CP, Swanson DB. Assessment of clinical skills with standardized patients: state of the art. Teach Learn Med. 1990;2:58–72.
16. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med. 1975;1:447–51.

V. Pascual-Ramos et al. / Reumatol Clin. 2018;**14(3)**:137–141   141

17. Raj N, Badcock LJ, Brown GA, Deighton CM, O'Reilly SC. Design and validation of 2 objective structured clinical examination stations to assess core under-graduate examination skills of the hand and knee. J Rheumatol. 2007;34: 421–4.
18. Hassell AB, West Midlands Rheumatology Services and Training Commit-tee. Assessment of specialist registrars in rheumatology: experience of an objective structured clinical examination (OSCE). Rheumatology. 2002;41: 1323–8.
19. Ryan S, Stevenson K, Hassell AB. Assessment of clinical nurse specialists in rheu-matology using an OSCE. Musculoskelet Care. 2007;5:119–29.
20. Criscione-Schreiber LG, Solana RJ, Hawley J, Jonas BL, O'Rourke K, Bolster MB. Expert panel consensus on assessment checklists for a rheumatology objective structured clinical examination. Arthritis Care Res. 2015;67:898–904.
21. Carraccio C, Englander R. The objective structured clinical examination: a step in the direction of competency-based evaluation. Arch Pediatr Adolesc Med. 2000;154:736–41.
22. Mc Gaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evi-dence. Acad Med. 2011;86:706–11.