

Evaluation of Clinical Competence in Trainers of Family and Community Medicine Residents

M. Pedregal González,^a E. Molina Fernández,^a J.A. Prados Castillejo,^b F. Quesada Jiménez,^c P. Bonal Pitz,^d and C. Iglesias Arrabal^e

Objectives. We show the first experience of the application of an objective and structured clinical evaluation (OSCE) procedure to family medicine trainers, that has been carried out in Andalucía. The objective is to use a competence evaluation instrument that, in the short term, will be used not only for trainer accreditation but also for other public sanitary professionals.

Participants. Tutors of family and community medicine residents.

Design. Observational descriptive.

Setting. Educational unity of family medicine.

Principal measurements. The competencial components to be assessed are the following: anamnesis, physical exploration, communication, technical skill, management, family attention and preventive activities.

The clinical situations were selected using the following priority criteria: prevalence, clinical gravity, prevention and early diagnosis importance, case complexity, doctor's capacity of evaluation and simplicity.

Results. Thirteen family medicine trainers took part in the OSCE. Their average age was 42.8±3.6 years. The test had an overall reliability coefficient (Cronbach's alpha) of 0.73. The overall mean score of the participants was 73±6.2. The best results about the competencial components were family attention, communication and technical skill.

Conclusions. The OSCE can be a convenient tool for family medical trainer evaluation, helping to orientate their education in the weak points and, in the near future, it can also be used as an instrument to do accreditation family medicine trainers.

Key words: Clinical competence. Family medicine trainer. Clinical evaluation.

EVALUACIÓN DE LA COMPETENCIA CLÍNICA DE TUTORES DE RESIDENTES DE MEDICINA FAMILIAR Y COMUNITARIA

Objetivo. Describir la primera experiencia de una evaluación clínica objetiva y estructurada (ECOFE) a tutores de residentes realizada en la comunidad autónoma de Andalucía.

Diseño. Observacional descriptivo.

Emplazamiento. Unidad Docente de Medicina Familiar y Comunitaria de Huelva.

Participantes. Tutores de residentes de medicina familiar y comunitaria.

Mediciones principales. Los componentes competenciales que se consensuaron y ponderaron fueron los siguientes: anamnesis, exploración física, comunicación, habilidades técnicas, manejo, atención a la familia y actividades preventivas.

La selección de las 10 situaciones clínicas de las que constaba la prueba se realizó utilizando unos criterios de priorización según la prevalencia, gravedad clínica, importancia de la prevención y del diagnóstico precoz, complejidad del caso, evaluación de la capacidad resolutoria del médico y simplicidad evaluativa.

Resultados. Realizaron la ECOFE 13 tutores. La edad media ± desviación estándar de los participantes fue de 42,8 ± 3,64 años. La prueba tuvo un coeficiente de fiabilidad (alfa de Cronbach) de 0,73. Por lo que se refiere a los resultados por participantes, la media global fue de 73 ± 6,2. Al analizar los resultados para los diferentes componentes competenciales, los mejores resultados se obtuvieron en la atención a la familia, la comunicación y las habilidades técnicas.

Conclusiones. Las pruebas de evaluación clínica objetiva y estructurada pueden ser útiles para la valoración de los tutores, con el objetivo de orientar su formación en los puntos débiles e incluso para, en un futuro no muy lejano, servir como instrumento para acreditar y reacreditar a los tutores de residentes de medicina familiar y comunitaria.

Palabras clave: Competencia clínica. Tutores de residentes de medicina de familia. Evaluación clínica.

Spanish version available at
www.atencionprimaria.com/62.820

A commentary follow
 this article
 (pág. 73)

Specialists in Family and
 Community Medicine

^aUnidad Docente de Medicina Familiar y Comunitaria de Huelva, Spain.

^bCentro de Salud Polígono de Guadalquivir, Córdoba, Spain.

^cCentro de Salud de Cartuja-Almanjáyar, Granada, Spain.

^dUnidad Docente de Medicina Familiar y Comunitaria, Centro de Salud Universitario de Bellavista, Seville, Spain.

^eCasa del Mar (Instituto Social de la Marina), El Puerto de Santa María, Cádiz, Spain.

Financial support of the Juan Ramón Jiménez Hospital.

Correspondence: Dr. Miguel Pedregal González. Unidad Docente de Medicina Familiar y Comunitaria de Huelva. Hospital Juan Ramón Jiménez. Ronda Norte, s/n. 21005 Huelva. España.

Manuscript received 24 October, 2003.
 Manuscript accepted for publication 7 January, 2004.

Introduction

Developed over 20 years ago, objective structured clinical examination (OSCE) is an assessment method that is used by a growing number of health institutions, medical schools and scientific societies. Its use has spread throughout nearly all English-speaking countries.¹

In Spain, specifically in Catalonia, the Institute of Health Studies of the Catalanian Regional Government (Generalitat) has carried out assessment exercises since 1994 to evaluate graduate and postgraduate trainees in different medical specialties.^{2,3} This has made it possible to detect gaps in health care skills among both students and practitioners, and to identify changes needed in medical training activities.⁴

The number of clinical cases that should be included in an OSCE reflects a compromise between validity, reliability and feasibility. The use of 10 or more cases ensures good reliability, but in general most OSCE include approximately 20 cases distributed in 30 stations.⁵

The specialty designated family and community medicine comprises a wide range of knowledge fields and skills, and family physicians have pioneered efforts to implement innovative training methods such as a specific program for this specialty, continuing education, certification for training activities, and postgraduate training and evaluation.⁶

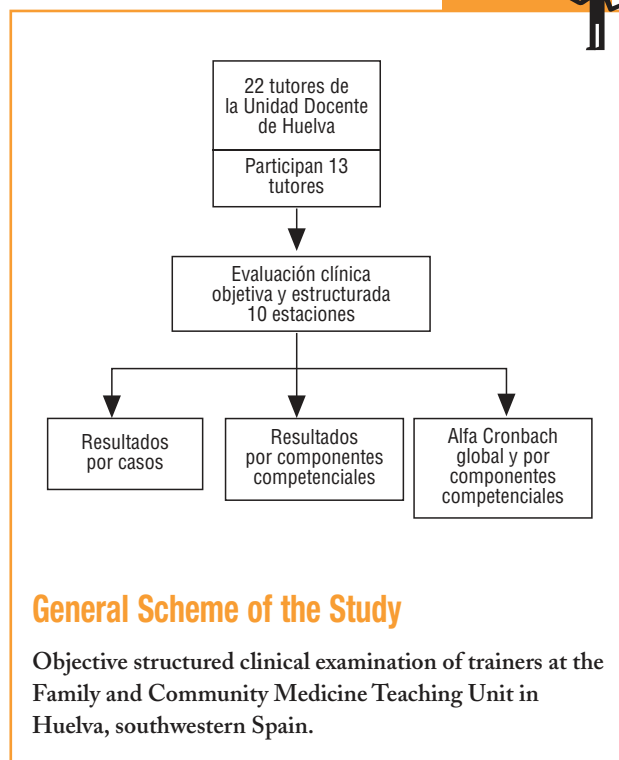
Among the criteria proposed for certification and recertification for trainers of family and community medicine residents is the trainer's professional competence. The OSCE has been proposed as a reliable way to evaluate this.⁷

Few assessment exercises have been carried out in Spain with the OSCE instrument to evaluate trainers of family and community medicine residents, and no such assessments have been attempted in the autonomous community of Andalusia, the region comprising most of southern Spain. We report the first such experience, carried out at the Family and Community Medicine Teaching Unit in the city of Huelva, southwestern Spain. The aim of the exercise was to test the usefulness of the OSCE in evaluating clinical skills and detecting areas in need of improvement, to enable trainers to enhance their own training.

Material and Methods

The test committee was constituted in 2001. The committee consisted of 6 family physicians experienced in clinical practice at primary care centers, and in the training of residents as tutors and course leaders. All committee members had knowledge of and experience in the evaluation of competence. The functions of the test committee were to define components of com-

Material and methods



petence, prepare the list of clinical situations that merited evaluation, construct the table of specific components of competence, and design some of the clinical cases to be used in the evaluation.

The components of competence and weightings, chosen by consensus, were based on recommendations of the Catalanian Society of Family and Community Medicine,⁸ and were weighted as follows: clinical interview, 12%; physical examination, 12%; interpersonal communication, 16%; technical skill, 10%; management (ability of the participant to reach a correct diagnosis from the information obtained from the clinical interview, physical examination, and complementary tests; knowledge of treatments; ability to devise a treatment plan), 40%; attention to family members, 4%, and preventive activities, 6%.

The choice of clinical situations was agreed by consensus on the basis of the following priority criteria: prevalence, clinical severity, importance of prevention and early diagnosis, complexity of the case, evaluation of the physician's ability to reach a decision, and feasibility of evaluation.⁹ The most appropriate instrument for evaluation was chosen depending on the clinical situation: standardized patient, mannequin, clinical image interpretation, set of open questions requiring brief answers, or multiple-choice questions. The table of specific components contained the competence areas to be evaluated and their corresponding clinical situation and instrument. Two clinical cases with their script, practical instructions and scoring checklist (along with additional advice) were supplied by the Institute of Health Studies. All other clinical cases were developed by the test committee.

Logistics and psychometric analyses were managed by the staff of the Family and Community Medicine Teaching Unit in Huelva.

TABLE 1
Stations and Instruments Used

Station No.	Name of Case	Type of Station
1	Back pain	Standardized patient
2	Diabetes	Multiple choice questions
3	Chest pain	Standardized patient
4	Basic CPR	Mannequin
5	COLD	Standard patient
6	Spot on skin	Image interpretation + open questions
7	Depression	Standardized patient
8	Palpitations	Image interpretation + open questions
9	Headache	Standardized patient
10	Minor surgery	Mannequin

The test consisted of 10 clinical situations, and candidates to be evaluated were allowed 6 minutes per station. This period reflects the mean duration of each patient-physician contact at primary care centers in the region of Andalusia. The stations represented a variety of situations commonly encountered in primary care: emergencies, scheduled appointments and minor surgery (Table 1). The exercise was divided into 2 groups of stations (rounds), and no breaks were allowed between stations. The total time allowed to complete the test was 104 minutes.

The OSCE was carried out at the La Orden health center in the city of Huelva on May 16, 2003. This center was chosen because it provided an appropriate physical environment for the test.

Participants were chosen from among trainers of residents at the Family and Community Medicine Teaching Unit in Huelva, and participation was voluntary. Other participants in the exercise were 5 standardized patients, 7 observers and 3 persons responsible for logistics.

The number of stations was set at 10 on the basis of available resources, and because this was the minimum number of stations that yields reliable results.

The highest score possible for each case was 100 points, distributed between the various components of competence to be evaluated. The total score was calculated as the mean of the scores for each case. At the end of the exercise all participants were given a questionnaire to solicit their opinions and feelings about the exercise.

The data were analyzed with SPSS software (version 11.0). Reliability of the test in general and of each component was calculated as Cronbach's alpha. Correlations of the scores with age and with years as a trainer were calculated as Pearson's correlation coefficient. The results per component of competence were found for each participant, and mean scores and standard deviations were also calculated. The scores per case were also calculated, as were the results for each participant.

Results

Thirteen of the 22 certified tutors at the teaching unit (8 men, 5 women) participated in the OSCE. Mean age was 42.8 years (SD, 3.6 years), and mean duration of

TABLE 2
Cronbach's Alpha for Each Component of Competence

Competence	Cronbach's alpha
Preventive activities	0.7921
Interpersonal communication	0.6453
Technical skill	0.6277
Attention to family	0.5637
Clinical interview	0.5443
Physical examination	0.4007
Management	0.2987
Global	0.7287

teaching and tutoring activities for residents at the time of the test was 8.6 years (SD, 4.6 years). Of the 9 trainers who did not take part in the OSCE (4 men, 5 women), 8 had personal reasons (the exercise was held on a Friday afternoon) and one woman was on maternity leave. Mean age of the 9 nonparticipants was 41.2 years (SD, 2.4 years) and mean time as a trainer and tutor was 4.4 years (SD, 2.9 years). There were no significant differences in age between participants and nonparticipants in the OSCE ($P=.27$), although the difference in the number of years as a trainer and tutor was significant ($P=.03$).

The global reliability coefficient for the test (Cronbach's alpha) was 0.7287. By component of competence, the highest scores were obtained for preventive activities, interpersonal communication and technical skill (Table 2). The global mean score for all participants was 73 (SD, 6.2), with a range of 64.4 to 81.9 and an interquartile range of 11.7 (Figure 1).

The best scores were obtained for the components of competence attention to family members, technical skill and interpersonal communication (Table 3). The worst scores

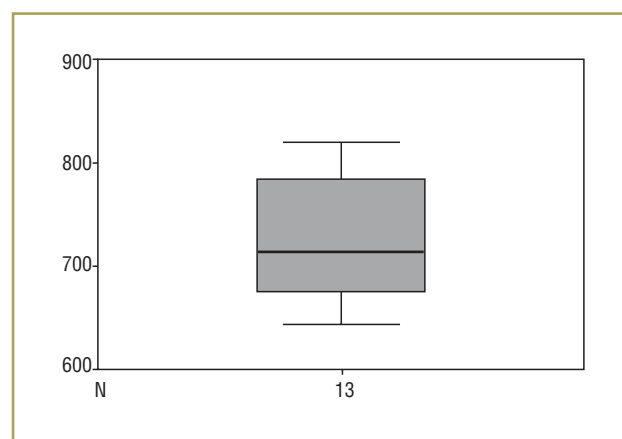
**FIGURE 1**
Distribution of trainers' scores by case.

TABLE 3 Results According to Component of Competence

Competence	Mean	SD
Preventive activities	61.53	29.95
Physical examination	65.76	10.96
Clinical interview	70.44	16.76
Management	71.61	10.11
Interpersonal communication	80.71	7.68
Technical skill	81.07	15.21
Attention to family	84.23	23.01

TABLE 4 Results According to Case

Case No.	Mean	SD
1	59.77	7.25
2	92.62	14.03
3	74.00	11.69
4	56.54	28.53
5	43.15	25.61
6	73.85	13.40
7	80.15	10.84
8	78.46	22.30
9	85.38	10.50
10	84.54	14.08

were for preventive activities and physical examination, including clinical interview and management. Table 4 shows the mean score and standard deviation for each case. Test scores and ages of the trainer showed a normal distribution (Shapiro-Wilk's test). Correlations were sought between scores and years as a trainer or age with Pearson's correlation coefficient. Neither correlation was significant. Most participants (10 of 13) agreed partially or completely that the test and the time allowed for each station reflected conditions encountered in actual practice. Approval was also expressed with the items on the participants' evaluation form that inquired about how the stations were designed, interpretation of the cases, and whether this type of evaluation was useful in improving training for residents. All participants agreed that this type of exercise measured professional capacity better than classical written examinations. The open item on the evaluation form provided space for participants to express emotional features of the test, and elicited comments such as "I was really nervous," "I felt really stressed and a bit nervous," "My self-esteem has hit bottom," "I really liked it," "Overall, I felt comfortable," and "I really enjoyed it."

Discussion

Reliability is the best documented psychometric parameter in evaluations with simulated patients. This indicator is defined as a measure of the reproducibility of a process across a series of replications.¹⁰ A minimum of 10 cases are needed to achieve good reliability. In the present study, Cronbach's alpha was similar to values reported in earlier research.² The global findings for components of competence were satisfactory. However, it was surprising that the skill that yielded the lowest mean score was preventive activities (notwithstanding the large standard deviation), a result that reflected the wide variation in individual scores. This component of competence thus requires improvement. The second lowest mean score was obtained for physical examination skills. This may reflect a loss of skills with time, probably because the patient load leads to simplification of some steps in the physical examination. On the other hand, the highest scores were obtained for attention to the patient's family (although the weighting of this element in the overall OCSE score was low), interpersonal communication, and technical skill. These results reflect the conscientiousness with which practitioners treat

Discussion
Key points



What Is Known About the Subject

- Developed more than 20 years ago, objective structured clinical examination is a method of assessment that is used by a growing number of health institutions.
- The Institute of Health Studies of the Catalanian Regional Government (Generalitat) has carried out assessment exercises since 1994 for graduate and postgraduate training in different medical specialties.
- Few assessment exercises have been carried out with objective structured clinical examination to evaluate the competence of trainers of family and community medicine residents.

What This Study Contributes

- The evaluation method yielded a Cronbach's alpha of 0.7287, a result consistent with earlier studies.
- Mean global score for participants in this study was 73 (SD, 6.17), with scores ranging from 64.4 to 81.9.
- The best results for specific components of competences were obtained for attention to family members, communication and technical skill.

these elements, and possibly the efforts the teaching unit devotes to training the trainers in these areas.

Of the different situations encountered in the exercise, the lowest scores and greatest dispersion were obtained for cases 4 (basic CPR) and 5 (chronic obstructive pulmonary disease). The results for case 4 may reflect the fact that only 4 of the trainers who performed the OSCE worked in rural health centers where a physician is on duty around the clock. This requires them to keep their training in basic life support techniques up to date. In case 5 the greatest weighting was given to the preventive activities component, according to which physicians should give the patient brief counseling about the benefits of quitting smoking. Overlooking this element lowered the score for this case considerably. The stress of the evaluation exercise, and the short time allowed for each case (although the period was a reflection of the time available in real-life conditions) might explain this result. In any case we believe the global outcome and scores for the remaining cases can be considered good.

In general, participants were willing to accept the exercise although some trainers were nervous, and one participant noted that the test had affected self-esteem.

Our findings show that the teaching unit for family and community medicine should make efforts to improve the components of competence that yielded the worst results. Further training will be offered in preventive activities such as counseling and other elements of the motivational interview. Training in physical examination skills will also be improved with appropriate tools, including workshops, CDs, standardized patients, and videotapes of encounters with real patients. These measures will probably lead to improvements in residents' training.

Objective structured clinical examinations can be useful for evaluating trainers with the aim of enhancing their training in areas where skills are weak. In the not-too-distant future OSCE exercises may serve as a certification and recertification tool for trainers of residents in family and community medicine.

Acknowledgments

We thank the test committee, the trainers who volunteered to participate in this study, Milagros Villamarín Casal for her dedication, Juan Pedro Moreno García and the Huelva-Costa Health District for their assistance, and José Luis Sánchez Ramos for his judicious advice, and the

References

1. Blay Pueyo C. Evaluación clínica objetiva y estructurada: ECOE. *Educación Médica* 1998;1:13-6.
2. Ruiz i Gil E, Florensa i Claramunt E, Cots i Yago JM, Sellares Salas J, Iruela López A, Blay Pueyo C, et al. Primeras experiencias en evaluación de la competencia clínica de los médicos de familia de Catalunya. *Aten Primaria* 2001;28:105-9.
3. Cervera Alemany AM, Salvà Casanovas A, Altimir Losada S, Miralle Basseda R, Yuste Marco A, Blay Pueyo C, et al. Evaluación clínica objetiva y estructurada en geriatría. Resultados del primer proyecto realizado en España. *Rev Esp Geriatr Gerontol* 2002;37:298-303.
4. Martínez Carretero JM, Blay Pueyo C. Evaluando la competencia de los profesionales sanitarios: dudas fundamentos y experiencias. *Educación Médica* 1998;1:17-23.
5. Colliver JA, Vu NV, Barrows H. Screening test length for sequential testing for a SP examination: a receiver operating characteristic (ROC) analysis. *Acad Med* 1992;67:592-5.
6. Sellares Sallas J, Cots Yago JM. ¿Debemos evaluar la competencia de los médicos de familia? *Aten Primaria* 2002;30:539-40.
7. Gómez Gascón T. Acreditación y reacreditación de tutores de MFyC en España. *Aten Primaria* 2002;29:164-6.
8. Blay C, Cordón F, Cots JM, Ezquerro M, Florensa E, Forés D, et al. Certificació-recertificació de la competencia profesional dels metges d'atenció primària. *Butlletí Societat Catalana de Medicina Familiar i Comunitària* 1996;63:9-16.
9. Sellarés J, Bosch C, Florensa E, De Serdio E, Gorroñoigoitia A, Medina E, et al. Evaluación de la competencia: reto y necesidad. Barcelona: semFYC, 2002.
10. Colliver J, Williams R. Technical issues: test application. *Acad Med* 1993;68:454-60.

COMMENTARY

The Limits of Objective Structured Clinical Examination

C. Blay

Médico de Familia, ICS ABS Santa Eugènia de Berga, Berga, España, Unitat Avaluació Competència Clínica, Institut d'Estudis de la Salut, Barcelona, Spain.

As predicted in the 1990s,¹ the competence of health care practitioners and ways in which competence can be evaluated have become current topics of debate within medical circles. The number of institutional, legislative, and research initiatives that have appeared show clearly that the debate on professionalism that has been taking place throughout the world has not been ignored in our milieu. Consequently, projects in Spain that set out to evaluate physicians' clinical competence are becoming increasingly numerous and visible.

The excellent article this editorial is based on is a perfect example of this phenomenon, in two ways. On one hand it reflects the methodological rigor that should be required of any exercise intended to evaluate professional competence. In addition, it reveals the limits, paradoxes and areas of progress that typify evaluation—issues that I will discuss briefly in the following sections.

Establishing Strategic Foundations

Any evaluation exercise should establish its strategic foundations clearly. This means that the project should specify *who* is to be evaluated, what *level of performance* will be required, and especially, what the *purpose* of the evaluation is. These characteristics will determine the structure and content of the exercise.

Evaluation of the complex construct known as *professional competence* requires different formats depending on the ideology that underlies each project². Different formats are appropriate for different situations, and evaluation with an objective structured clinical examination (OSCE) is not suitable for all purposes.

Formative Versus Summative Evaluation

These characteristics of OSCE are particularly evident when the choice is between evaluation aimed at certifying competence (summative) and evaluation aimed at guiding training (formative).

When summative evaluation is required, the reliability of the test, its predictive validity and construct validity are absolutely crucial. When a formative approach is needed,

Key Points

- A current topic in debates on professionalism is the spread of models for evaluating professional competence.
- Objective structured clinical examination has been shown to be an excellent method for certifying competence, although its capacity to generate formative feedback is limited.
- Poor evaluations are worse than no evaluations at all.

verification efforts should center on the format's educational impact and content validity.

An OSCE comprising only 10 stations may be a good instrument for summative certification, but regardless of its reliability, it is of limited use in providing high-quality formative feedback, if quality is understood to mean that the feedback is true, complete and useful to the subject.

Interpreting Psychometric Parameters Appropriately

One of the risks of professional competence evaluation is the trivialization of statistical considerations. Reliability (understood as internal consistency and measured with Cronbach's alpha) is often at the heart of disputes over the usefulness of a given evaluation as a measuring instrument. Although reliability is a crucial consideration in any exercise designed for certification, a few brief comments are in order.

1. Sometimes—especially in exercises with a small number of participants—the alpha value is artificially high. When this happens an apparently reliable test can lead to unfair decisions.

2. Validity is as important as reliability. Drawing conclusions—especially for summative evaluations—before the prototype test has been *validated*, is surely ill-advised and unfair to the professionals who are to be evaluated.

3. When the aim is to provide feedback to persons who are being evaluated regarding their competence profile, it must first be determined to what extent each of the components of competence behaves reliably. This is generally not accomplished with small-scale OSCE.

What OSCE Are Good for

Objective structured clinical examination is one test format among many potentially appropriate combinations of testing instruments. Although OSCE is the paradigm for evaluating practical clinical skills in a simulated environment, it is not a structure that is suitable for all purposes. The OSCE is an excellent way to certify professional competence, and especially to detect individuals whose skills are clearly inadequate. It is also capable of providing some measure of feedback on the skills of the collective that participate in the exercise.

But in general, its capacity to provide formative feedback to individuals is limited to the global results (“What’s my global score in comparison to others in the group?”). Thus an OSCE is inadequate for characterizing an individual’s profile as an actual practitioner. As a mechanism of formative feedback, reflection on “how I felt while I was taking the OSCE” is a much more powerful instrument than “what I scored on each component of competence.”

Competence of Expert Professionals

Competence is a dynamic phenomenon³ that is changeable and can be expressed in different ways as the profes-

sional acquires experience. Two experienced professionals may both be competent and make equally correct decisions when faced with particular situations, but may use completely different approaches.

Moreover, professional competence goes beyond clinical dexterity. Elements of the affective and moral world, and aspects related with professional attitude, are even *more* important components of competence in expert professionals. This is what makes evaluating experienced professionals a much more complex process than evaluating junior colleagues. Examining more experienced practitioners thus requires a larger and more refined combination of evaluation tools.

Evaluating professionals makes sense only if the results of the exercise are true, fair and useful in terms of decisions making, self-reflection and motivation to improve. Nothing could be more contrary to this end than the inappropriate use of evaluation tools.

In view of the growing presence of such tools, we should make efforts to put them to good use—for the good of the profession.

References

1. Blay C. Evaluación de la competencia profesional: ¿están cambiando los tiempos? *Aten Primaria* 1995;16:2-4.
2. Newble D, Jolly B, Wakeford R. The certification and recertification of doctors: issues in the assessment of clinical competence. Cambridge: Cambridge University Press, 1994.
3. Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA* 2002;287:226-35.