



## ORIGINAL

# La rúbrica en el examen oral de Traumatología y Ortopedia



Plácido Zamora Navas\* y Enrique Guerado Parra

Área de Traumatología y Cirugía Ortopédica, Departamento Especialidades Quirúrgicas, Facultad de Medicina, Málaga, España

Recibido el 7 de marzo de 2017; aceptado el 7 de abril de 2017

Disponible en Internet el 7 de junio de 2017

### PALABRAS CLAVE

Rúbrica;  
Examen oral;  
Educación médica  
grado;  
Evaluación sumativa;  
Evaluación

### Resumen

**Objetivo:** Conocer si el uso de la rúbrica para el examen oral de Traumatología y Ortopedia aporta información sobre competencias transversales.

**Tipo de estudio:** Experimental con un grupo control.

**Material y métodos:** Grupo control: 67 alumnos calificados globalmente mediante examen oral. Grupo intervención: 68 alumnos valorados con rúbrica que contiene las dimensiones: conocimiento teórico, precisión en la terminología, concreción en el lenguaje y uso de fuentes adicionales de información.

**Resultados:** Existe diferencia entre las calificaciones del grupo control (mediana 9, RIC: 9, 9) y el experimental (mediana 7,63, RIC: 6,48, 8,54). En el grupo experimental, la calificación más alta se obtiene en la pregunta propuesta por el profesor y la más baja en la elegida por el alumno. Precisión y concreción han alcanzado un 72 y un 70% del máximo posible. El conocimiento teórico ha conseguido el 76,26%. Se observa un efecto monotonía de las calificaciones del grupo control a partir del alumno evaluado en décimo lugar.

**Conclusiones:** Utilizar rúbrica influye en la calificación final del alumno. La precisión y la concreción en la exposición alcanzan un mínimo del 70% de la valoración adscrita.

© 2017 Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\* Autor para correspondencia.

Correo electrónico: [pzn@uma.es](mailto:pzn@uma.es) (P. Zamora Navas).

**KEYWORDS**

Rubric;  
 Oral exam;  
 Undergraduate  
 medical education;  
 Summative  
 evaluation;  
 Assessment

**The rubric in the oral exam in Traumatology and Orthopaedics****Abstract**

*Aim:* To find out if the use of the rubric for the oral examination of Traumatology and orthopaedics provides information on cross-disciplinary skills.

*Study type:* Experimental with a control group.

*Material and methods:* Control Group: 67 students globally qualified by oral examination. Intervention Group: 68 students evaluated with a scoring rubric that comprises the dimensions: Theoretical knowledge, precision in the terminology, accuracy in the language, and the use of additional information sources.

*Results:* There was a significant difference between the results of the control group (median 9, IQR: 9.9) and the experimental group (median 7.63, 95% CI: 6.48 - 8.54). In the experimental group, the higher qualification is obtained in the question proposed by the teacher and the lower one in the question chosen by the student. Precision and specificity were 72% and 70% of the maximum, respectively. Theoretical knowledge achieved 76.26%. A monotony effect is observed in the scores of the control group after the tenth evaluated student.

*Conclusions:* The use of a rubric influences the final assessment of the student. The precision and accuracy achieved a minimum of 70% of the assigned valuation.

© 2017 Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Introducción**

La evaluación es una parte sustantiva del proceso de enseñanza-aprendizaje. Para este fin, la prueba más extendida es el examen tipo test multirrespuesta que no está libre de matices en cuanto al planteamiento de las preguntas y la interpretación de estas<sup>1</sup>.

El examen oral puede abarcar otras consideraciones que la evaluación del conocimiento teórico, a la vez que da opción a la interacción entre examinador y examinado para esclarecer dificultades en la exposición y profundización en temas. Así, el examen oral está considerado la forma de evaluación más completa en la taxonomía de Bloom<sup>2</sup>. De hecho, permite evaluar otras dimensiones como el razonamiento clínico, el tiempo empleado en estudiar o la forma de comunicar las ideas<sup>3</sup>. Además, es reconocida como una oportunidad de aprendizaje por el propio alumno<sup>4</sup>.

Las mayores dificultades de este tipo de evaluación tienen que ver con la falta de confiabilidad y de validez<sup>5</sup>.

Para explotar correctamente las ventajas del examen oral, este debe diseñarse previamente como también la ponderación a otorgar a otros aspectos. Para controlar la subjetividad es necesario que queden recogidos en un documento, la rúbrica, y publicados con anterioridad.

**Hipótesis de investigación**

La introducción de la rúbrica en el examen oral de Traumatología y Ortopedia mejora la objetivación del proceso de aprendizaje sin minusvalorar la calificación del conocimiento.

**Objetivo principal**

Conocer la información adicional que la rúbrica supone en la evaluación del examen oral de Traumatología y Ortopedia.

**Objetivos específicos**

- Saber si el uso de rúbrica supone un detrimento en la valoración del conocimiento teórico.
- Conocer el nivel de precisión en la terminología.
- Analizar el grado de concreción en el uso del lenguaje.
- Estimar el uso de fuentes adicionales.
- Valorar si el uso de una rúbrica durante un examen oral es de utilidad en la objetivación de este.

**Material y métodos****Tipo de estudio**

Estudio experimental con grupo control.

**Ámbito de estudio**

Alumnos matriculados en la asignatura de Enfermedades del Aparato Locomotor del 5.º año del grado de Medicina.

**Grupos de estudio**

- Control: alumnos evaluados mediante prueba oral. Evaluación: 0 a 10. Aprobado en 5.
- Intervención: alumnos evaluados mediante prueba oral con rúbrica. Evaluación: 0 a 10. Aprobado en 5.

**Aleatorización**

Correlativamente, sorteando el alumno por el que iniciar la prueba. Posteriormente y en grupo de 3, asignación a un grupo u otro de forma consecutiva.

## Cálculo del tamaño muestral

Para el objetivo principal. Potencia del 80%. Diferencias en el contraste de la hipótesis nula para  $H_0: \mu_1 = \mu_2$  mediante prueba de la t de Student bilateral para muestras independientes. Nivel de significación 5%. Asumiendo media del grupo control de 6,77 puntos y de 5,77 el de intervención. Desviación típica de ambos: 2,11 unidades; 61 individuos en el grupo control y 63 en el de intervención. Por posibles pérdidas, se incrementó un 10%.

## Desarrollo

Han participado 2 profesores. Cada uno ha evaluado de forma independiente y en su totalidad a uno de los grupos.

Grupo control: mediante batería de preguntas a criterio del profesor dentro del programa. Grupo intervención: utilización de rúbrica estructurada con 3 preguntas:

1. Nuclear. Elegida por el profesor y considerada esencial. Valoración: 50% del examen.
2. Propuesta por el alumno. Valoración: 30% del total.
3. Adicional. A elección del profesor. De tipo general y conocimiento accesorio. Valor: 20%. En cada pregunta se valoran 4 dimensiones con 5 grados de exhaustividad.

Dimensiones: contenido. Valor: 75% del total de la pregunta cuando es respondida en el mayor grado de exhaustividad.

1. Precisión. Valor: 20%
2. Concreción. Calificación: 3%.
3. Información adicional. Total: 2%.

Exhaustividad:

1. No respuesta.
2. Faltan bastantes contenidos importantes.
3. Falta algún contenido importante.

4. Presente el contenido importante, falta algún contenido no básico.
5. Incluye todo el contenido.

Para responder al último objetivo específico, se ha utilizado examen tipo test multirrespuesta con una sola correcta y cebo de 0,33.

## Análisis estadístico

Análisis descriptivo. Las variables continuas se resumieron en medias y desviación estándar o medianas según la distribución. Rango de valores: máximo y mínimo.

Para analizar las diferencias entre variables cuantitativas continuas, prueba de la t de Student para 2 muestras independientes en caso de normalidad, comprobado mediante el test de Shapiro-Wilk. En caso de no normalidad, prueba no paramétrica U de Mann-Whitney y prueba consistente de Welch.

## Resultados

### Descripción de la muestra

La muestra la componen 135 alumnos (67 en el grupo de control y 68 en el de intervención) (tabla 1).

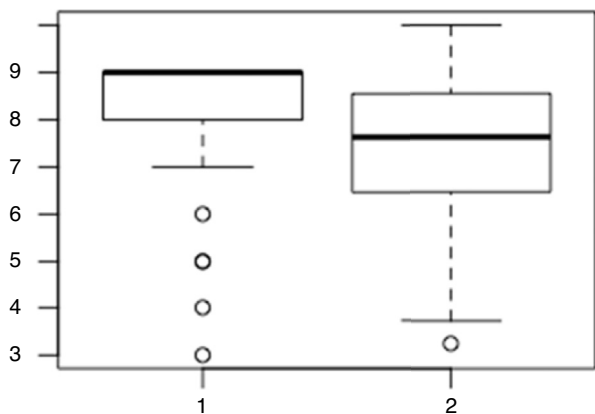
La distribución de la variable principal, nota media (grupos control, intervención), no presentó una distribución normal, Shapiro-Wilk ( $W=0,94353$ ,  $p=0,003979$ ;  $W=0,58516$ ,  $p=1,81E-012$ , respectivamente) (fig. 1).

### Contraste

Para saber si el uso de rúbrica supuso un detrimento en la valoración se ha realizado una comparación con la calificación alcanzada en el examen oral frente a la obtenida en el examen tipo test. La diferencia fue estadísticamente significativa ( $t=5,1572$  [Welch],  $df=128,59$ ,  $p$  valor =  $9,228E-07$ ).

**Tabla 1** Descripción de la muestra

	Grupo control		Grupo intervención	
Hombres		31		26
Mujeres		36		42
	Ex oral	Test	Rúbrica	Test
Nota máxima	9	10	10	9,8
Nota mínima	3	2,4	3,25	3,4
Suspensos	2	14	5	12
Sobresalientes	47	3	16	2
Media	8,3	6,5	7,48	6,2
Desviación estándar	1,3	1,7	1,59	1,4
Curtosis	5,9	-0,2	-0,12	-0,2
Coefficiente asimetría	-2,4	-0,6	-0,57	-0,2
Mediana	9	6,6	7,63	6,3
1 cuartil	8	5,5	6,48	5,5
2 cuartil	9	6,6	7,63	6,3
3 cuartil	9	7,6	8,54	7,3
4 cuartil	9	10	10	9,8



**Figura 1** Caja y bigotes de las características de las calificaciones en los grupos.

Se contrastó si la calificación obtenida por los alumnos en los grupos control y de intervención fue diferente. Esta diferencia es estadísticamente significativa ( $t = -3,3117$  [Welch],  $df = 126,24$ ,  $p = 0,00121$ ; IC  $[-1,347193, 0,339374]$ ).

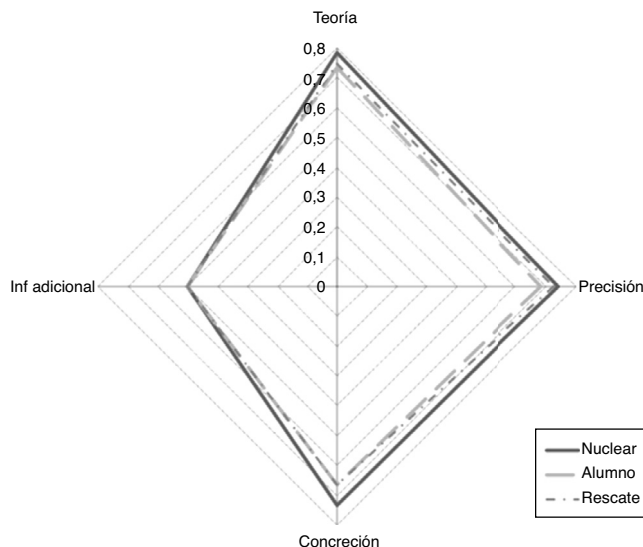
Dentro del grupo de intervención, se deseó conocer, de no haber considerado otras dimensiones, cuál habría sido la calificación final obtenida. Una vez ponderadas las calificaciones, considerando solo la dimensión teoría, habría sido de 7,75 (RIC: 6.45, 8.75). Esta diferencia no se demostró estadísticamente significativa ( $W = -2.139,5$ ,  $p = 0,4538$ ).

**Análisis de la rúbrica (tabla 2)**

**Dimensión «conocimiento teoría»**

El valor asignado al conjunto de las 3 preguntas fue del 75% de la calificación total. La calificación ponderada habría sido de 7,60 (RIC: 6,51, 7,67). Esta diferencia no ha sido significativa en su comparación con la calificación original ( $t = -1,1014$  [Welch],  $df = 92713$ ,  $p = 0,2736$  [IC:  $-0,9510108, 0,2724394$ ]). Tampoco ha sido diferente la comparación con la valoración del examen sin rúbrica ( $t = -1,7887$  [Welch],  $df = 76,631$ ,  $p = 0,07762$ ). También se ha comparado la nota obtenida en el test de conocimientos teóricos con la obtenida en la dimensión de teoría del examen oral realizado con rúbrica, no encontrándose significación estadística ( $t = -1,9445$  [Welch],  $df = 124,7$ ,  $p$  valor =  $0,05409$ ).

La calificación del conocimiento teórico tuvo un valor sumatorio de las 3 preguntas de 5,72, con un aprovechamiento del 76,26% del máximo asignado. Para la pregunta



**Figura 2** Resumen del peso de las diferentes dimensiones en el total de la evaluación.

propuesta por el profesor alcanzó un 78% del valor total asignado, el 75% para la pregunta adicional y el 73% para aquella que fue elegida por el alumno.

**Conjunto otras dimensiones**

Para este apartado se propuso un 25% de la calificación global; el 20% para el concepto precisión, el 3% para concreción y el 2% para uso de fuentes adicionales. Finalmente, este total añadido tuvo un valor máximo de 2,5 puntos sobre 10 y un mínimo de 0,65 con una mediana de 1,45, lo que supone un peso del 15% de la calificación global con un aprovechamiento de oportunidad de un 58%.

**Precisión.** El nivel de precisión fue de 1,44 (RIC 1,2-1,65) con un aprovechamiento del 72% del peso decidido previamente, un total del 19,34 de la calificación final absoluta. El nivel más bajo correspondió a la pregunta elegida por el alumno (68,33%) y el mejor en la pregunta nuclear (74%).

**Concreción.** Alcanzó una mediana de 0,2175 (RIC 0,1875-0,24) y un peso de 0,0286 del porcentaje de la nota final. Se aprovechó un 70% de la oportunidad. La concreción fue mayor en la pregunta formulada por el profesor (73%). Las otras 2 preguntas alcanzaron el 66,66%.

**Uso de fuentes adicionales.** Con un peso del 2% de la calificación global, llegó al 0,1 en las 3 preguntas (RIC: 0,063-0,13), un aprovechamiento del 50% de la oportunidad. No se observó diferencias entre las 3 preguntas (fig. 2).

**Tabla 2** Valores alcanzados en cada una de las preguntas y sus diferentes dimensiones

	Teoría (7,5 p)	Precisión (2 p)	Concreción (0,3 p)	Inf. adic. (0,2 p)	Total	% aprov.
Nuclear (5 p)	2,4	0,7	0,1	0,05	3,8	76,8
Alumno (3 p)	1,6	0,4	0,1	0,03	2,1	71,7
Rescate (2 p)	1,1	0,3	0,04	0,02	1,5	74
Total	5,7	1,4	0,2	0,1		
% aprov.	76,3	72	70	50		

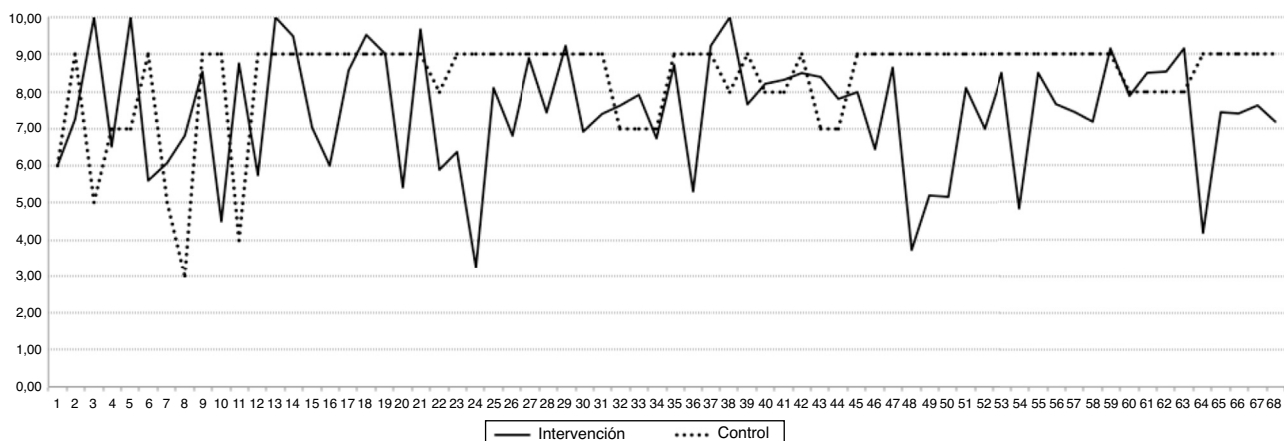


Figura 3 Evolución temporal de las calificaciones.

**Otros hallazgos.** Se realizó una valoración de la evolución de las calificaciones conforme avanzaba la prueba. Se encontró un efecto modulación en la variación en las calificaciones del grupo control a partir del alumno evaluado en décima posición. Los 9 primeros obtuvieron una mediana de 6,5 (RIC: 5, 7,5) y los posteriores 9 (RIC: 9, 9). En el grupo de intervención, no se encontró este efecto.

Cuando se comparan las calificaciones de los 2 subgrupos del grupo control con las obtenidas en el grupo de intervención, no se encuentra significación estadística en la comparación de los alumnos evaluados al principio de la prueba (hasta el alumno 10) ( $t = 1,2982$  [Welch],  $df = 9,3509$ ,  $p$  valor =  $0,2253$ ), y sí en la comparación entre el segundo subgrupo del grupo control y el de intervención ( $t = -4,2242$  [Welch],  $df = 104,67$ ,  $p$  valor =  $5,132E-05$ ) (fig. 3).

## Discusión

La capacidad del examen oral para la interacción y la empatía con el alumno ha hecho que algunos autores la consideren como la prueba ideal para evaluar a alumnos que precisen una valoración personalizada; singularmente aquellos que opten a calificación brillante, o los considerados «en el límite»<sup>6</sup>.

El uso de la rúbrica puede ayudar en su aplicación<sup>7</sup>. Aunque el empleo de esta tiene una gran implantación en la actualidad<sup>8</sup>, su utilización comenzó a extenderse a finales del siglo xx<sup>9</sup>.

La rúbrica es considerada por los alumnos como una excelente guía de orientación formativa. No obstante, sus declaradas ventajas, y aunque entre el profesorado también se reconoce como una herramienta útil para dirigir la evaluación formativa<sup>10</sup>, existe una resistencia para su uso entre los evaluadores<sup>11</sup>. En ocasiones, esta se debe a una falta de reconocimiento de sus ventajas<sup>12</sup> y en otras por el innegable consumo de tiempo en la preparación y por precisar de formación específica<sup>13</sup>.

El uso de la rúbrica tiene como reto mejorar la validez y la reproducibilidad de las pruebas orales. Si bien la reproducibilidad alcanza niveles aceptables (70%)<sup>14</sup>, su validez es

más difícil de apreciar debido a los múltiples aspectos que se deben considerar (constructo, apariencia, etc.)<sup>15</sup>.

Es esencial definir con anterioridad el peso de las diferentes dimensiones a considerar; conocimiento teórico y otras competencias transversales. De esta manera, mientras el profesor establece los criterios de evaluación de una forma objetiva, el alumno obtiene una guía para diseñar el autoaprendizaje<sup>16</sup>.

La rúbrica debe ser diseñada con criterios de coherencia con el objetivo a evaluar, apropiada para el nivel que se examina y ser diseñada con anterioridad<sup>17</sup>.

La rúbrica empleada en este estudio ha sido de tipo analítico, mediante la cual se ha pretendido conocer las dimensiones en las que los alumnos han basado la estructuración de su conocimiento y en qué medida se ha profundizado.

En el presente estudio se ha utilizado el contraste de las calificaciones obtenidas mediante examen tipo test. La calificación alcanzada en el examen oral se ha encontrado estadísticamente significativa y en términos absolutos superior en 1,34 puntos. No obstante, cuando se han comparado la calificación ponderada del componente «conocimiento de la teoría» con la prueba escrita, la diferencia no se ha encontrado significativa. Esto nos permite inferir que la discrepancia en calificaciones se debe a una cuantificación adicional de las competencias transversales definidas.

En la exploración de las competencias transversales, los alumnos han alcanzado valores que se sitúan en un aprovechamiento de la oportunidad del 70-75% de la puntuación máxima asignada.

Una vez ponderada la calificación obtenida en las 3 preguntas, la calificación alcanzada ha sido similar, un 70% del máximo, no encontrándose diferencia en la calificación ponderada de las 3 preguntas.

La evaluación mediante rúbrica para valorar competencias transversales ha demostrado su correlación con el examen oral estructurado, aunque tiene poca relación con la apreciación del conocimiento teórico<sup>18</sup>. En nuestra investigación no existe correlación con la evaluación del conocimiento teórico pero existe diferencia



estadísticamente significativa con la calificación final obtenida. Esto implica a la par que la calificación global del alumno no se ve minorada con la utilización de una rúbrica y que es capaz de premiar la incorporación de otras habilidades.

Algunos autores consideran que no existe correlación entre las calificaciones tras examen escrito u oral<sup>19</sup>. Nosotros hemos encontrado una diferencia entre las calificaciones de la prueba escrita y la oral, no así entre la prueba escrita y su comparación con la medida exclusivamente de la dimensión teórica del examen oral.

Se ha recomendado realizar examen tipo test como complemento al oral por no ser capaz este de valorar correctamente la dimensión de conocimiento teórico<sup>20</sup>. Lejos de considerarse excluyentes, examen oral y escrito pueden ser entendidos como complementarios, aunque la calificación de los contenidos teóricos se ha reflejado correctamente con la evaluación del examen oral atendiendo a la elaboración previa de una rúbrica, en algunos alumnos pueden encontrarse un efecto techo.

### Hallazgos incidentales

Las pruebas orales deben ser sometidas a un escrutinio para mejorar su validez, reproducibilidad y justicia<sup>15</sup>. De entre las exigencias para las pruebas orales, las condiciones del examinador son reconocidas como una de las 6 dimensiones a considerar<sup>21</sup>. Si bien el propósito de excelencia es reconocido en el evaluador, este puede verse afectado por circunstancias como el efecto saturación<sup>22</sup>.

Hemos encontrado un efecto modulación en las calificaciones dentro del grupo control. En este, las puntuaciones desde el alumno evaluado en posición décima en adelante han presentado pocas variaciones, mientras que en la serie intervención se ha mantenido la dinámica anterior.

La exigencia de que todos los alumnos puedan beneficiarse de las ventajas que aporta la prueba oral exige que todos sean evaluados en las mismas condiciones. La necesidad de tener que usar un tiempo limitado para evaluar a un número alto de alumnos es una dificultad añadida. Esto solo es solucionable con la programación a lo largo de varias clases, consumiendo así un tiempo del que generalmente el alumno no dispone<sup>23</sup>.

### Conclusiones

El empleo de rúbrica en el examen oral enriquece significativamente la valoración de otras dimensiones del proceso de aprendizaje.

La precisión y la concreción alcanzan valores de un mínimo del 70% de la valoración adscrita.

El uso de fuentes de información adicional es la dimensión más débil dentro de la calificación final.

Ni la selección de preguntas de obligado conocimiento, ni de otras ya sean consideradas accesorias o libremente elegidas por el alumno, consiguen una valoración diferencial.

### Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

### Bibliografía

1. Wiggins G. A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*. 1989;70:703-13.
2. Bloom B. Taxonomy of educational objectives: The classification of educational goals: Handbook 1: Cognitive domain. New York: NY: David MacKay; 1971.
3. Allen D, Tanner K. Rubrics: Tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE Life Sciences Education*. 2006;5:197-203.
4. Centeno A, Primogero C. Do students learn during oral exams? Keywords: Evaluation, learning. *Assessment and evaluation. Educación Médica*. 2004;7:187.
5. Davis MH, Karunathilake I. The place of the oral examination in today's assessment systems. *Med Teach*. 2005;27:294-7.
6. Ghosh A, Mandal A, Das N, Tripathy SK, Biswas A, Bera T. Students' performance in written and viva-voce components of final summative pharmacology examination in MBBS curriculum: A critical insight. *Indian J Pharmacol*. 2012;44:274-5.
7. Jonsson A, Svingby G. The use of scoring rubrics: Reliability, validity and educational consequences. *Educ Res Rev*. 2007;2:130-44.
8. Lynd-Balta E. Using literature and innovative assessments to ignite interest and cultivate critical thinking skills in an undergraduate neuroscience course. *CBE Life Sci Educ*. 2006;5:167-74.
9. Luft JA. Rubrics: Design and use in science teacher education. *J Sci Teacher Educ*. 1999;10:107-21.
10. Schneider JF. Rubrics for teacher education in community college. *Community College Enterprise*. 2006;12:39-55.
11. Parkes KA. The effect of performance rubrics on college-level applied studio grading. PhD diss, University of Miami. UMI No. 3215237, 2006.
12. Green R, Bowser M. Observations from the field: Sharing a literature review rubric. *J Libr Admin*. 2006;45:185-202.
13. Hafner JC, Hafner PM. Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *Intl J Sci Educ*. 2003;25:1509-28.
14. Dunbar NE, Brooks CF, Kubicka-Miller T. Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. *Innovat High Educ*. 2006;31:115-28.
15. Memon MA, Joughin GR, Memon B. Oral assessment and post-graduate medical examinations: Establishing conditions for validity, reliability and fairness. *Adv in Health Sci Educ*. 2010;15:277.
16. Hyungkyu K, Giljae L, Eunjung L. Associate profesor. *Educación Médica*. 2004;7:211.
17. Gática-Lara F, Uribarren-Berrueta T. ¿Cómo elaborar una rúbrica? *Inv Ed Med*. 2013;2:61-5.
18. Westkämper R, Hofer R, Weber M, Aeschlimann A, Beyeler C. Is it possible to conduct high-stake oral examinations in a reliable and valid way for small numbers of candidates with limited resources? *Educación Médica*. 2004;7:159.
19. Torke S, Reem RA, Ramnarayan K, Asha K. The impact of vivavoce examination on students' performance in theory component of the final summative examination in physiology. *J Physiol Pathophysiol*. 2010;1:10-2.
20. Gortney JS, Lundquist LM. Experience and evaluation of oral examinations in a therapeutics course for second-year pharmacy students. *Curr Pharm Teach Learn*. 2013;5:321-30.
21. Joughin G. Dimensions of oral assessment and student approaches to learning. En: Brown S, Glasner A, editores. *Assessment matters*. Buckingham: The Society for Research into Higher Education & Open University Press; 1999. p. 146-56.

22. American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 1999.
23. Boedigheimer R, Ghrist M, Peterson D, Kallemyn B. Individual oral exams in mathematics courses: 10 years of experience at the Air Force Academy. PRIMUS. 2015;25: 99–120.