

Evaluación de competencias transversales por rúbricas. Aplicación a competencia de presentación mediante elementos multimedia

J. Orozco-Messana^a, A. Navarro Gomez^b y M. Siurana Paula^c

^aDepartamento de Ingeniería Mecánica y de Materiales (jaormes@upv.es), ^bInstituto de Ciencia y Tecnología del Hormigón (alnagme@upvnet.upv.es) y ^cDepartamento de Matemática Aplicada (masiupau@upv.es).

Abstract

The need to evaluate transversal competencies in teaching has grown without applying a systematic research to the procedures used. On this paper we present a rubric based procedure for evaluating the presentation competency from a video based item. This paper validates a procedure allowing independent neutral evaluations and providing results statistically independent from other conflicting elements thanks to the multimedia support.

Keywords:

Transversal competencies, evaluation, video rubric

Resumen

La necesidad de evaluar competencias transversales ha crecido en los últimos años sin disponer de una investigación sistemática sobre los procedimientos a utilizar. Este trabajo valida un procedimiento que permite obtener resultados neutrales e independientes en base al tratamiento estadístico de los resultados sin variables cruzadas a través del apoyo multimedia.

Palabras clave:

Competencias transversales, evaluación, rúbrica para videos

Introducción

Este trabajo analiza la validez de la evaluación sumativa de las habilidades de comunicación oral con apoyo visual de videos. Se parte del establecimiento de una rúbrica genérica calificando la validez de su aplicación a través de la convergencia de los resultados obtenidos por distintos profesores (con diferentes trasfondos) en disciplinas muy diferenciadas. Se parte de la presentación de trabajos de grupos de estudiantes universitarios para su evaluación por profesorado de diversas áreas. La evaluación se centra en la competencia transversal de comunicación oral, sin interferencia con la calidad académica de los contenidos presentados.

En los últimos años se ha evidenciado en nuestro contexto universitario un notable auge de los modelos de enseñanza centrados en el aprendizaje, en gran parte promovidos por el Espacio Europeo de Educación Superior, que enfatizan la importancia de integrar los procesos de enseñanza, aprendizaje y evaluación en el desarrollo de una amplia variedad de tareas abiertas, auténticas y realistas -p.e., proyectos de trabajo, realización de investigaciones, análisis de casos, ...- que faciliten la adquisición de las competencias generales y específicas consideradas en las distintas titulaciones.

Desde esta perspectiva resulta ineludible considerar técnicas de evaluación alternativas, no convencionales o innovadoras, que permitan (a) utilizar la evaluación como un recurso instruccional más para promover el aprendizaje activo, la puesta en marcha de procesos cognitivos de alto nivel y la adquisición de competencias, (b) facilitar a los estudiantes los criterios de ejecución a alcanzar en la realización de las tareas, favoreciendo el desarrollo de las habilidades de autorregulación y autoevaluación realista de sus trabajos, (c) recibir retroalimentación específica sobre cómo mejorar sus niveles de ejecución y, (d) facilitar al profesorado información de los resultados de aprendizaje que van alcanzando sus estudiantes. Las estrategias de evaluación tradicionales resultan claramente insuficientes para conseguir estos propósitos, dado que (a) los resultados en los exámenes tradicionales pueden ayudar en parte a los estudiantes a monitorizar su aprendizaje, pero muy poco a promoverlo y, (b) plantear tareas auténticas y realistas implica considerar un amplio rango de respuestas y elaboraciones satisfactorias posibles que difícilmente pueden quedar reflejadas en las mismas (Huba y Freed, 2000).

Las rúbricas constituyen herramientas de evaluación no convencionales que pueden definirse como guías para evaluar la calidad de las elaboraciones y el nivel de ejecución alcanzado por los estudiantes en una amplia variedad de tareas complejas, especificando los criterios a considerar y los niveles de adecuación en cada uno de ellos (desde inadecuado a excelente) (Andrade y Du, 2005). En la actualidad, las rúbricas constituyen un método de evaluación extendido en la enseñanza obligatoria, especialmente en el ámbito anglosajón véase la amplitud de recursos web disponibles en Dornish y Sabatini (2006)-, aunque su consideración todavía resulta escasa en nuestro contexto.

Quizá una de las cuestiones que explica en mayor grado las resistencias del profesorado universitario a introducir este tipo de herramientas de evaluación sea la controversia sobre su fiabilidad y validez para valorar las elaboraciones de los estudiantes –algunos autores sugieren que junto a su escasa formación docente y sobre técnicas de evaluación-, cuestión que debe seguir ocupando un lugar central en la investigación (Malini y Andrade, 2010). Desde esta perspectiva, a nadie le resultará extraño que numerosos estudios previos se hayan centrado en (a) analizar la fiabilidad intrajueces e interjueces en su aplicación para valorar los trabajos de los estudiantes (p.e., Thaler, Kazemi y Huscher, 2009), (b) la consistencia entre las valoraciones efectuadas por los estudiantes al evaluar el trabajo de sus compañeros o sus propias elaboraciones (p.e., Roblyer y Wiencke, 2003; Sadler y Good, 2006) y, (c) la consistencia entre los criterios que utilizan estudiantes y profesorado en su aplicación (p.e., Hafner & Hafner, 2003; Kocakula, 2010; López-Pastor, Fernández-Balboa, Santos y Fraile, 2011; Stellmack, Konheim-Kalskstein, Manor, Massey y Schmitz, 2009).

Sus conclusiones básicas son que la rúbricas permiten incrementar la consistencia entre las valoraciones interjueces e intrajueces, facilitando al profesorado emitir juicios válidos sobre el nivel de ejecución de los estudiantes al evaluar competencias complejas (Jonsson y Svingby, 2007).

En definitiva, se detecta una necesidad de verificar la validez de la utilización de rúbricas y se desarrolla un trabajo experimental que avale los procedimientos habituales a través de la consistencia en la aplicación de una rúbrica tipo en la evaluación la competencia transversal de comunicación oral, por parte del profesorado de distintas áreas. Esto se consigue estableciendo un análisis de la relación entre las calificaciones derivadas de las mismas. Adicionalmente, también se analiza la percepción de utilidad y validez que los rúbricas pueden tener como soporte para el desarrollo de sus presentaciones orales con apoyo visual.

El trabajo experimental parte de una muestra de áreas tecnológicas muy diferentes a fin de sistematizar la aplicación de la rúbrica seleccionada para evaluar la competencia de presentación en público verificando la objetividad de los resultados a través del análisis estadístico cruzado entre profesores.

1. Procedimiento utilizado

1.1. Método

Se analiza el nivel de convergencia entre las valoraciones efectuadas por 3 profesores de distintas áreas, y sobre 6 presentaciones de alumnos, desde una perspectiva analítica – criterio a criterio- y holística – puntuación global-. La percepción de validez y utilidad se determina a partir de las valoraciones efectuadas por los profesores en relación a la rúbrica propuesta y que se basa en los usos habituales en las titulaciones de la UPV.

Siguiendo los principios destacados en la introducción, se ha seleccionado la rúbrica desarrollada por la ETSID (de la UPV) para a evaluar las habilidades de presentación oral con apoyo visual, dado que es la más frecuentemente utilizada para evaluar esta competencia transversal en las distintas Escuelas de la UPV, durante la evaluación de los Proyectos Final de Grado (PFGs) se especifican los siguientes elementos:

- Criterios de evaluación. Incorpora once criterios distintos en la evaluación de las elaboraciones de los estudiantes, aplicados sobre el documento soporte (PPT) de la presentación en sí mismo y la fase de presentación (centradas en el discurso y su resumen).
- Niveles de ejecución. La rúbrica considera cinco niveles de ejecución graduados en función de su adecuación para cada presentación.
- Descripción de los niveles de ejecución. Los niveles de ejecución se describen de forma clara, reflejando las diferencias entre los mismos.
- Estrategia de calificación. La rúbrica incorpora una valoración cuantitativa para cada criterio y nivel de ejecución, permitiendo efectuar una valoración analítica (criterio a criterio, asignando a cada uno de ellos una valoración entre 0 y 5) y holística (sumatorio de las puntuaciones obtenidas en todos los criterios, con un rango entre 0 y 42) de las presentaciones efectuadas por los estudiantes.

En el estudio participan 9 estudiantes de las titulaciones de: Máster Universitario en Tecnologías Industriales (2), Grado en Ingeniería de Obras Públicas (2) y Grado en Biotecnología (5), en el curso académico 2016-2017. Un 44% son mujeres y un 56% varones. Las edades oscilan entre los 21 y 24 años y todos son estudiantes a tiempo completo, mientras que el 83% desarrolla algún tipo de actividad profesional. La distribución por sexo, edad y dedicación es característica en este tipo de estudios universitarios en la Universitat Politècnica de Valencia.

Tras especificar a los estudiantes el proyecto de trabajo a desarrollar (elaboración y presentación oral con apoyo visual con grabación en video) se presentaron y discutieron los objetivos académicos de la competencia transversal de presentación oral, sin concretar las líneas específicas de evaluación en la rúbrica.

Con el objetivo de familiarizar a los estudiantes en su aplicación, facilitar su comprensión y eliminar posibles resistencias, se efectuó y discutió su aplicación para evaluar dos presentaciones previas efectuadas por el propio profesor. Seis grupos de trabajo desarrollaron y efectuaron sus presentaciones, que fueron evaluadas a través de la aplicación de la rúbrica por 3 profesores de distintas áreas de conocimiento (los firmantes del artículo). La aplicación de la rúbrica se cumplimentó individualmente tras observar los videos de cada presentación.

1.2. Percepción de utilidad y de validez

Siguiendo los principios y orientaciones de la investigación previa (p.e., Moskal, 2003; Moskal y Leydens, 2000), se desarrolló un cuestionario dirigido a evaluar la percepción de utilidad de la rúbrica para el desarrollo del proyecto de trabajo (10 ítems) y la percepción de validez de la misma (11 ítems). Se utilizó una escala de respuesta tipo Likert de cinco niveles (desde “totalmente inadecuado” valorado con 1, hasta “totalmente adecuado” con un 5).

Los profesores evaluaron la calidad de las presentación oral a través de la rúbrica ([www.etsid.upv.es/.../Borrador%20Trabajo%20Fin%20de%20Gradov11%20\(002\).pdf](http://www.etsid.upv.es/.../Borrador%20Trabajo%20Fin%20de%20Gradov11%20(002).pdf)) más consensuada en la UPV, tanto desde un punto de vista analítico como holístico. La valoración analítica permitirá determinar el nivel de acuerdo criterio a criterio entre profesores. La valoración holística permitirá determinar la existencia de una relación significativa entre las valoraciones globales efectuadas por ambos. La aplicación de la rúbrica se hizo en diferido sobre los videos grabados de las presentaciones.

La percepción de la utilidad y validez de la rúbrica se determina a través de la concordancia estadística de las valoraciones de los 3 profesores participantes pertenecientes a diversas áreas y con bagajes muy diferentes.

El grado de acuerdo en la aplicación de la rúbrica se calcula con dos procedimientos de estimación distintos, uno conservador y otro laxo, aceptados y comúnmente utilizados en este ámbito (Tinsley y Weiss, 2000). En el procedimiento conservador se define que existe acuerdo entre profesores cuando otorgan exactamente la misma valoración (nivel) en los

criterios de la rúbrica. El procedimiento laxo considera que existe acuerdo entre profesores cuando otorgan una valoración en los criterios que difiere como máximo un nivel entre los mismos. En ambos casos, el grado de acuerdo o convergencia se determina a través del estadístico kappa. También se calculó el nivel de asociación entre las valoraciones globales efectuadas por los profesores (sumatorio de las puntuaciones en los distintos criterios, con un máximo de 55 puntos) a través del coeficiente de correlación de Pearson.

1.3. Resultados

En la tabla 1 se recogen las medias y desviaciones típicas de las respuestas independientes de cada elemento de valoración de la rúbrica propuesta, así como el coeficiente Kappa de Fleiss aplicada a una comparativa entre alumnos.

Posteriormente, la tabla 2 realiza un análisis similar pero para la valoración entre distintas presentaciones.

Se puede observar que las valoraciones obtenidas tiene baja dispersión y alta Kappa de Fleiss ($\kappa > 0,7$). El nivel de acuerdo entre profesores es alto (κ siempre superior a 0,7) en cualquier tipo de valoración. Las diferencias del análisis comparativo y el laxo son bajas correspondiendo a una valoración fuerte de la concordancia.

**Tabla 1. Resultados del análisis comparativo
de valoraciones para las 6 presentaciones (9 alumnos)**

Dimensiones de la escala	Media	Desviación típica	K de Fleiss
1	3,9	0,5	0,72
2	4,2	0,6	0,84
3	4,2	0,4	0,94
4	3,4	1,5	0,70
5	4,3	0,5	0,98
6	3,7	0,4	0,82
7	4,1	0,5	0,83
8	1,3	0,9	0,99
9	3,8	0,7	0,75
10	4,1	0,6	0,88
11	4,3	0,8	0,96

Tabla 2. Resultados del análisis laxo de la rúbrica por presentación (6 presentaciones)

Proyecto presentado	Puntuación total	Acuerdos en valoración	K de Fleiss
1	39,0	54	0,85
2	40,9	62	0,92
3	39,3	66	0,94
4	40,1	66	0,91
5	44,0	64	0,88
6	43,4	64	0,81

1.4. Discusión y conclusiones

El primer objetivo del trabajo ha consistido en analizar la percepción de validez de la rúbrica presentada en este trabajo. Los resultados son excelentes. En definitiva, la rúbrica es de utilidad para promover y dar soporte a procesos cognitivos de alto nivel, favoreciendo la concordancia en la evaluación de la competencia transversal de presentación mediante herramientas multimedia. La rúbrica propuesta es adecuada pues para evaluar la competencia transversal de presentación en este tipo de proyectos, destacando que integra los criterios y elementos clave a considerar en su elaboración, que permite evaluar competencias importantes para el ejercicio profesional y que les ha permitido desarrollar habilidades que utilizarán en situaciones futuras.

Estas conclusiones coinciden, básicamente, con las destacadas en estudios previos, centrados en otras competencias y titulaciones universitarias (p.e., Cothran, 2003; Hafner & Hafner, 2003; Rohrbach, 2008; Struyven, Dochy & Janssen, 2008). Sin embargo, a partir de las valoraciones obtenidas, cabría plantearse en trabajos futuros (a) reducir el número de criterios que incorpora, coincidiendo con el principio de que un número reducido de criterios resulta más práctico para el desarrollo de las actividades de aprendizaje (p.e., Thaler et al., 2009) y, (b) analizar la moderada percepción de utilidad para reducir la ansiedad que genera la elaboración de los proyectos de trabajo -quizá provocada porque las presentaciones orales siguen constituyendo una de las tareas en que los estudiantes universitarios manifiestan mayores niveles de ansiedad (García-Ros & Pérez-González, 2011b)-, que sí se destaca en estudios previos (p.e., Andrade & Du, 2005).

Desde una perspectiva analítica (análisis criterio a criterio), el grado de acuerdo entre profesores es significativo, mostrando un grado de convergencia incluso superior a los escasos estudios previos dirigidos a determinar la utilidad de las rúbricas como soporte para la evaluación a través de los pares de las elaboraciones de los estudiantes (p.e., Sadler & Good, 2006; Stellmack et al., 2009).

Desde una perspectiva holística, la relación entre la valoración global de las presentaciones efectuada por los profesores alcanzó una relación de elevada magnitud. De forma más concreta, la mayoría de trabajos que analizan esta cuestión, Falchicov & Goldfinch (2000), obtienen valores similares.

Estos resultados apoyan la idea de que los profesores aplican la rúbrica de forma similar y que puede constituir una herramienta válida para valorar y calificar las presentaciones de los estudiantes, facilitando adicionalmente la comprensión de los criterios de calidad implicados en los proyectos de trabajo (Lu & Law, 2011). Todos estos aspectos destacan la importancia de seguir centrandó la investigación en determinar las medidas que pueden permitir incrementar la validez y fiabilidad de este tipo de herramientas como soporte para la autoevaluación y heteroevaluación de las elaboraciones de los compañeros (Malini & Andrade, 2010).

1.5. Referencias complementarias

Andrade, H. (2001). Using rubrics to promote thinking and learning. *Educational Leadership*, 57 (5), 13-18. Andrade, H. (2005). Teaching with rubrics: The Good, the Bad, and the Ugly. *College teaching*, 53 (1), 27-31.

Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*, 10 (3). Recuperado el 11 de mayo de 2009, <http://pareonline.net/getvn.asp?v=10&n=3>. Andrade, H.; Du, Y., & Mycek, K. (2010).

Baron, J., & Keller, M. (2003). Use on rubrics in online assessment. *Evaluation and Assessment Conference*, 23-25 novembre 2003.

Cho, K.; Schunn, Ch. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, 98(4), 891-901. Cothran, D. J. (2003).

Students's use of and perspective on rubrics. *Educational Research, Risks and Dilemmas. NZARE/AARE Conference*. Auckland, Nueva Zelanda. Recuperado el 17 de octubre de 2009, <http://www.aare.edu.au/03pap/cot03119.pdf>.

Dornisch, M., & Sabatini, A. (2006). Limitations of web-based rubric resources: Addressing the challenges. *Practical Assessment Research & Evaluation*, 11(3). Recuperado el 11 de mayo de 2010, <http://pareonline.net/getvn.asp?v=11&n=3>.

Falchicov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A metaanalysis comparing peer and teacher marks. *Review of Educational Research*, 70 (3), 287-322.

García-Ros, R., & Pérez-González (2011). Assessment preferences of preservice teachers: analysis according to academic level and relationship with learning styles and motivational orientation. *Teaching in Higher Education*, published online ahead of print publication, doi: 10.1080/13562517.2011.570434.

