

Análisis y validación de una rúbrica para evaluar habilidades de presentación oral en contextos universitarios

Rafael García-Ros¹

¹ Departamento de Psicología Evolutiva y de la Educación,
Universidad de Valencia

España

Correspondencia: Rafael García-Ros. Departamento Psicología Evolutiva y de la Educación. Avda. Blasco Ibáñez, 21. 46010-Valencia. España. E-mail: Rafael.Garcia@uv.es

© Education & Psychology I+D+i and Editorial EOS (Spain)

Resumen

Introducción. Este trabajo analiza la validez convergente entre la evaluación sumativa de las habilidades de comunicación oral con apoyo visual de estudiantes universitarios efectuada por el profesorado y por los pares a través de la aplicación de una rúbrica.

Método. Se analiza el nivel de convergencia entre las valoraciones efectuadas por el profesorado y por los pares 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 estudiantes sobre ambos aspectos en relación a la rúbrica desarrollada.

Resultados. Los estudiantes señalan que la rúbrica resulta útil para explicitar y clarificar los criterios de valoración, planificar el desarrollo de su trabajo y evaluar los productos resultantes. También destacan su validez para evaluar este tipo de proyectos, integrando los criterios clave a considerar en su desarrollo. El grado de acuerdo en su aplicación entre profesores y pares es significativo, tanto en desde una perspectiva analítica como holística, especialmente entre las valoraciones globales de los proyectos de trabajo –correlación de .89-.

Conclusión. Se destaca la percepción de utilidad y validez de la rúbrica para promover y dar soporte a procesos cognitivos de alto nivel en el desarrollo de este tipo de proyectos de trabajo, así como para su evaluación y para el desarrollo de competencias importantes en el ejercicio profesional. La rúbrica constituye una herramienta válida para valorar y calificar las presentaciones de los estudiantes a través de los pares. Estas conclusiones se discuten en términos de sus implicaciones y repercusiones instruccionales.

Palabras Clave: Rúbricas, enseñanza universitaria, fiabilidad y validez, percepción de los usuarios, evaluación a través de los pares, validez convergente, aprendizaje autorregulado.

Recibido: 10/10/11

Aceptación inicial: 10/10/11

Aceptación final: 04/11/11

Analysis and validation of a rubric to assess oral presentation skills in university contexts

Abstract

Introduction. The main objective of this study was to analyze users' perceptions and convergent validity of peer- and teacher summative assessment using a rubric of students' oral presentation skills in university context.

Method. Peer- and teacher-assessment convergence was analyzed from an analytical and holistic perspective. Students' perception of validity and usefulness was determined from a questionnaire developed ad-hoc for this study.

Results. Students perceive that rubric is useful to explain and clarify assessment criteria, planning the development of the projects and evaluating their results. They also highlight its validity, integrating the key criteria to consider in the development and presentation of their projects. The measures of agreement between peer- and teacher assessment using the rubric was significant, both analytical and holistic perspective -correlation of .89-.

Conclusion. Results show the perceived usefulness and validity of the rubric to promote and support high-level cognitive processes in the development of the projects. The rubric is a valid tool to assess and rate the presentation of the students through the peer. These findings are discussed in terms of its instructional implications.

Keywords: Rubrics, higher education, reliability and validity, user perceptions, peer assessment, convergent validity, selfregulated learning.

Received: 10/10/11

Initial acceptance: 10/10/11

Final acceptance: 04/11/11

Introducción

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 o plantillas de evaluación

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; Kocaküla, 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).

El interés de la investigación sobre las rúbricas también se ha centrado en analizar si su utilización promueve un aprendizaje de mayor calidad (p.e., Andrade y Du, 2005; Andrade, Du y Micek, 2010; Kocaküla, 2010; Popham, 1997) y si facilita la autoevaluación y heteroevaluación de las elaboraciones de los estudiantes (p.e., Andrade, 2001; Baron y Keller, 2003; Cho, Schunn y Wilson, 2006; Magin y Helmore, 2001; Schafer, Swanson, Bené y Newberry, 2001). Las conclusiones básicas de estos trabajos son que las rúbricas pueden promover un aprendizaje de mayor calidad al permitir focalizar la atención sobre los criterios de bondad en la elaboración de los proyectos de trabajo y que facilitan a los estudiantes una autoevaluación mucho más precisa de sus propios trabajos. También que permiten al profesorado ofrecer juicios más válidos sobre el nivel de adquisición de competencias complejas, así como proporcionar retroalimentación más específica y centrada sobre la ejecución de los estudiantes (Jonsson y Svingby, 2007). Sin embargo, diversos estudios también constatan que muchos estudiantes pueden no considerar la totalidad de criterios que incorporan (p.e., An-

drade y Du, 2005), que pueden mostrar reticencias para utilizarlas en la evaluación de las elaboraciones de sus compañeros (Norcini, 2003) o que señalan la importancia de implicar a los estudiantes en su desarrollo para facilitar su comprensión y aplicación (Huba y Freed, 2000; Stix, 1997; Taggart, Phifer, Nixon y Wood, 2001).

En esta misma línea, la investigación también se ha centrado en analizar las percepciones de los estudiantes (y también del profesorado) sobre los beneficios instruccionales y la utilidad de las rúbricas como herramientas de evaluación –características, adecuación, validez, utilidad...-, dado su papel determinante sobre cómo enfocan el aprendizaje y estudio (p.e., García-Ros & Pérez-González, 2011; Sander, 2005; Struyven, Dochy & Janssens, 2005). Sus conclusiones destacan que: (a) tanto estudiantes como profesores señalan su utilidad para explicitar y clarificar los criterios de evaluación –*transparencia*-, (b) los profesores indican que promueven el desarrollo de prácticas reflexivas y que les permiten obtener mayor información sobre la efectividad de las mismas, que facilitan ofrecer retroalimentación de mayor calidad y sirven como soporte para que los estudiantes autoevalúen sus elaboraciones, (c) provocan una mayor implicación de los estudiantes en el desarrollo de las tareas (Jonsson & Svingby, 2007; Schamberr & Mahoney, 2006). Sin embargo, distintos trabajos también constatan que los estudiantes pueden percibir las rúbricas más como una herramienta para satisfacer las demandas del profesorado que como una representación de los criterios y estándares de calidad a considerar en sus elaboraciones (Andrade & Du, 2005) o que pueden manifestar dudas sobre su utilidad para autoevaluar sus elaboraciones e interpretar mejor la retroalimentación recibida (Baron & Keller, 2003). Así, la eficacia instruccional de las rúbricas puede verse gravemente afectada si, por ejemplo, los estudiantes consideran que no contemplan los criterios clave para el desarrollo de la tarea, que no les resultan útiles para mejorar el resultado de su trabajo o que no permiten evaluar adecuadamente la calidad de sus elaboraciones.

Descripción de la rúbrica desarrollada

Siguiendo los principios destacados en la investigación previa sobre cómo incrementar su utilidad, adecuación y eficacia instruccional (p.e., Andrade, 2001, 2005; Jonassen, Peck & Wilson, 1999; Tierney y Marielle, 2004), en la rúbrica desarrollada en este trabajo dirigida a evaluar las habilidades de presentación oral con apoyo visual (véase Apéndice 1) se especifican los siguientes elementos:

- *Criterios de evaluación.* Incorpora catorce criterios distintos en la evaluación de las elaboraciones de los estudiantes (p.e., “Apoyo visual: relevancia y adecuación de las imágenes, grafos y esquemas utilizados”).
- *Niveles de ejecución.* La rúbrica considera cuatro niveles de ejecución graduados en función de su adecuación para cada criterio (inadecuado, básico/a mejorar, adecuado y muy adecuado).
- *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 3) 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.

Objetivos

De acuerdo con los planteamientos precedentes, el objetivo fundamental de este estudio consiste en el análisis y validación de una rúbrica dirigida a evaluar las habilidades de presentación oral con apoyo visual en contextos universitarios. De este modo, se analiza la consistencia en su aplicación entre el profesorado y los estudiantes al valorar las elaboraciones de sus propios compañeros, así como la relación entre las calificaciones derivadas de las mismas. Adicionalmente, también se analiza la percepción de utilidad y validez que los estudiantes conceden a la rúbrica como soporte para el desarrollo de sus presentaciones orales con apoyo visual.

Método

Participantes

En el estudio participan 64 estudiantes de la titulación de Psicopedagogía en el curso académico 2008-2009. Un 83% son mujeres y un 17% varones, con un rango de edad entre 22 y 49 años (media de 29.5 años y d.s. de 5.2). Un 17% 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 Universidad de Valencia.

Procedimiento

Tras especificar a los estudiantes el proyecto de trabajo a desarrollar en grupo (elaboración y presentación oral con apoyo visual de los principios y la aplicación en el aula de un modelo instruccional) se presentaron y discutieron los criterios destacados 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. Catorce grupos de trabajo desarrollaron y efectuaron sus presentaciones, que fueron evaluadas a través de la aplicación de la rúbrica por el profesor y por sus propios compañeros. El cuestionario sobre utilización y percepción de validez de la rúbrica se cumplimentó individualmente de forma anónima tras concluir todas las presentaciones.

Medidas

Se utilizó un cuestionario dirigido a evaluar la percepción de utilidad y validez de la rúbrica desarrollada en este trabajo que, junto a la aplicación de la misma en la evaluación de las presentaciones de los estudiantes, sirve para obtener las puntuaciones de los sujetos en las distintas variables consideradas en el estudio.

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” (1) hasta “totalmente adecuado” (5)-.

Valoración de las presentaciones. Profesor y estudiantes evaluaron la calidad de las presentaciones a través de la rúbrica, tanto desde un punto de vista analítico como holístico. La valoración analítica permitirá determinar el nivel de acuerdo criterio a criterio entre profesor y pares. La valoración holística permitirá determinar la existencia de una relación significativa entre las valoraciones globales efectuadas por ambos.

Análisis de datos

La percepción de la utilidad y validez de la rúbrica se determina a través de los descriptivos básicos de las respuestas ofrecidas por los estudiantes a los distintos ítems que integra el cuestionario elaborado al efecto.

El grado de acuerdo en la aplicación de la rúbrica entre profesor y estudiantes 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 profesor y pares cuando otorgan exactamente la misma valoración (nivel) en los criterios de la rúbrica. El procedimiento laxo considera que existe acuerdo entre profesor y pares 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 el profesor y por los pares (sumatorio de las puntuaciones en los distintos criterios, con un máximo de 42 puntos) a través del coeficiente de *correlación de Pearson*.

Resultados

Percepción de utilidad y validez de la rúbrica

En la tabla 1 se destacan los descriptivos básicos de las respuestas al cuestionario sobre percepción de utilidad de la rúbrica. Las valoraciones son superiores a la media teórica de la escala de respuesta (3.0) en todos los ítems, a excepción de “reducir mi ansiedad en la realización del trabajo”. Los estudiantes destacan especialmente su utilidad para “conocer mejor los criterios de valoración” (media de 4.4), “evaluar la adecuación del trabajo presentado” (media de 4.1), “desarrollar expectativas ajustadas sobre qué se demandaba” (media de 4.0) y “planificar la elaboración del trabajo y de la presentación” (media de 4.0).

Tabla 1. Descriptivos básicos sobre percepción de utilidad de la rúbrica

<i>Me ha resultado útil para ...</i>	Media	d.s.
1.- Desarrollar expectativas ajustadas sobre qué se me demandaba	4.0	(0.80)
2.- Planificar la elaboración del trabajo y su presentación	4.0	(0.83)
3.- Revisar lo que iba haciendo para ajustarlo a los criterios establecidos	3.7	(0.74)
4.- Evaluar la adecuación del trabajo presentado	4.1	(0.75)
5.- Guiar el desarrollo del trabajo	3.8	(0.79)
6.- Tomar decisiones en el seno del grupo sobre cómo desarrollarlo	3.8	(0.74)

<i>Creo que me ha permitido ...</i>	Media	d.s
1.- Mejorar el “producto” o resultado final de mi trabajo	3.9	(0.89)
2.- Facilitar el desarrollo del trabajo	3.5	(0.79)
3.- Disminuir mi “ansiedad” en la realización del trabajo	2.7	(0.98)
4.- Conocer mejor los criterios de valoración	4.4	(0.61)

Tal como se muestra en la tabla 2, los estudiantes valoran adecuadamente la validez de la rúbrica, especialmente respecto a “integrar los elementos clave en el desarrollo de la tarea” (media de 3.9), “desarrollar criterios que aplicaré en situaciones futuras” (media de 3.8) y “permite evaluar competencias importantes para el psicopedagogo” (media de 3.8).

Tabla 2. Descriptivos básicos sobre percepción de validez de la rúbrica

<i>Pienso que la rúbrica...</i>	Media	d.s.
1.- Integra los elementos clave en el desarrollo de la tarea efectuada	3.9	(0.81)
2.- Permite evaluar competencias importantes para el psicopedagogo	3.8	(0.72)
3.- Me ha permitido desarrollar criterios que aplicaré en situaciones futuras	3.8	(0.67)
4.- Es una herramienta fiable (mide adecuadamente la calidad del trabajo)	3.6	(0.83)
5.- Destaca con claridad los niveles considerados en cada criterio	3.6	(0.82)
6.- Incorpora criterios muy específicos y poco útiles en otras situaciones	2.6	(0.92)
7.- Facilita una comparación justa entre los trabajos de los grupos	3.5	(0.96)
8.- Presenta sesgos o estereotipos (género, edad, situación personal,...)	2.3	(0.70)
9.- Me ha facilitado adquirir los criterios de una adecuada ejecución	3.6	(0.78)
10.- Incorpora excesivos criterios	3.2	(0.84)
11.- Integra criterios excesivamente genéricos	2.6	(0.93)

Consistencia entre las valoraciones del profesor y de los estudiantes

El nivel de acuerdo entre profesor y pares utilizando el procedimiento de estimación conservador es del 66% ($kappa = .36, p < .001$), mientras que el nivel de acuerdo utilizando el procedimiento laxo es del 98% ($kappa = .80, p < .001$). El análisis de las diferencias en la valoración global otorgada a las presentaciones por profesores y pares (sumatorio de las puntuaciones en los distintos criterios en una escala 0-42), indica que un 70% de las mismas difieren en cuatro puntos o menos y que la discrepancia mayor es de 7 puntos.

Tras efectuar la ponderación de las valoraciones del profesor y de los pares en cada uno de los bloques de criterios y transformar los resultados a una escala entre 0 y 10, se observa una relación significativa ($r = .89, p < .001$) entre las calificaciones derivadas de las valoraciones del profesor (rango 5.6-9.2, media de 7.3 y desviación típica de 1.1) y de las efectuadas por los compañeros (rango 6.2-9.3, media de 7.8, desviación típica de 1.0). Las calificaciones asociadas a las valoraciones efectuadas por los pares resultan en general supe-

riores (promedio de 0.5 puntos) a las efectuadas por el profesor, con una diferencia mayor de un punto en cuatro grupos y siendo la diferencia máxima de 1.4 puntos.

Discusión y conclusiones

El primer objetivo del trabajo ha consistido en analizar la percepción de utilidad y validez de la rúbrica desarrollada en este trabajo. Los estudiantes destacan que les ha resultado útil para planificar, desarrollar y evaluar sus propias elaboraciones, especialmente para clarificar y conocer mejor los criterios de valoración de sus trabajos, desarrollar unas expectativas ajustadas sobre el nivel de ejecución a alcanzar, planificar el desarrollo de sus presentaciones y evaluar la adecuación de los productos resultantes. En definitiva, los estudiantes destacan la utilidad de la rúbrica para promover y dar soporte a procesos cognitivos de alto nivel, favoreciendo los procesos de autorregulación académica en el desarrollo de sus proyecto de trabajo (Zimmerman & Schunk, 2001). Adicionalmente, los estudiantes perciben la rúbrica como una herramienta válida y fiable para evaluar 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) .

El segundo objetivo se ha dirigido a analizar la validez convergente entre la aplicación de la rúbrica por el profesor y por los pares en la valoración de las presentaciones de sus compañeros, así como de la calificación obtenida a partir de las mismas. Esta cuestión resulta especialmente relevante dado que supone considerar la adecuación de la rúbrica para dar soporte

a la evaluación a través de los pares, ahorrar tiempo al profesorado en la corrección de las elaboraciones y disponer de un indicador del grado en que los estudiantes han interiorizado los criterios que integra. Por otro lado, las habilidades para juzgar las elaboraciones de los compañeros de forma crítica y objetiva constituyen competencias básicas en cualquier campo profesional (Magin & Helmore, 2001) que deben desarrollarse en las titulaciones universitarias.

Desde una perspectiva analítica (análisis criterio a criterio), el grado de acuerdo entre profesor y pares 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). De forma más específica, Jonsson & Svingby (2007) señalan que los valores del nivel de acuerdo exacto entre jueces se sitúa en el rango 55-75% -en este estudio es del 66%- , que el nivel de acuerdo entre niveles adyacentes se sitúa en el 90% -en este trabajo es del 98%- y que los trabajos que utilizan el índice kappa presentan valores entre .45-.75 –en este trabajo se obtienen valores significativos de .36 para el criterio exacto y de .80 para niveles adyacentes-.

Desde una perspectiva holística, la relación entre la valoración global de las presentaciones efectuada por el profesor y por los pares también alcanzó una relación de elevada magnitud. De forma más concreta, la mayoría de trabajos que analizan esta cuestión obtienen valores entre .55 y .75 –Falchicov & Goldfinch (2000) destacan en su estudio metaanalítico un valor promedio de .69-, mientras que en este trabajo se alcanza un valor de .89. Sin embargo, la calificación derivada de las valoraciones de los pares resulta ligeramente superior a la otorgada por el profesor (diferencia promedio de 0.5 puntos).

Estos resultados apoyan la idea de que los estudiantes aplican la rúbrica de forma similar a como lo efectúa el profesor y que puede constituir una herramienta válida para valorar y calificar las presentaciones de los estudiantes a través de los pares, facilitando adicionalmente la comprensión de los criterios de calidad implicados en los proyectos de trabajo (Lu & Law, 2011). Sin embargo, los resultados también destacan la tendencia de los estudiantes a sobrevalorar las elaboraciones de los compañeros frente a la valoración realizada por el profesor, cuestión también evidenciada en un estudio previo relacionado con las habilidades de presen-

tación oral (Magin & Helmore, 2001), aunque no en otros (p.e., Freeman, 1995; Hugues y Large, 1993). Implicar a los estudiantes en el desarrollo de las rúbricas, así como proporcionar información y retroalimentación adicional sobre su aplicación a los estudiantes que muestran mayores divergencias con la aplicación efectuada por el profesorado puede facilitar una mejor comprensión de los criterios considerados y mejorar la exactitud de sus valoraciones (Cho et al., 2006; Kocaküla, 2010; Lindblom-Yläne, Pihlajamäki y Kotkas, 2006). Todos estos aspectos destacan la importancia de seguir centrando 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).

Referencias

- 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). Rubric-referenced self-assessment and middle school students' writing. *Assessment in Education: Principles, Policy & Practice*, 17 (2), 199-214.
- 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 meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70 (3), 287-322.
- Freeman, M. (1995). Peer assessment by groups of group work. *Assessment and Evaluation in Higher Education*, 20 (3), 289-299.
- García-Ros, R., & Pérez-González (2011a). 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.

- García-Ros, R., & Pérez-González, F. (2011b). Validez predictiva e incremental de las habilidades de autorregulación académica en el éxito académico universitario. *Revista de Psicodidáctica*, 16 (2), 231-250.
- Hafner, J., & Hafner, P. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education*, 25(12), 1509-1528.
- Huba, M.E., & Freed, J.E. (2000). *Learner-Centered Assessment on College Campuses: Shifting the Focus from Teaching to Learning*. Needham Heights, Allyn & Bacon.
- Hugues, I., & Large, B. (1993). Staff and peer-group assessment of oral communication skills. *Higher Education*, 18 (3), 379-385.
- Jonassen, D., Peck, K., & Wilson, B. (1999). *Learning with technology*. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2 (2), 130-144.
- Kocaküla, M. S. (2010). Development and application of a Rubric for Evaluating Students' Performance on Newton' Laws of Motion. *Journal of Science Education Technology*, 19 (2), 146-164.
- Lindblom-Yläne, S.; Pihlajamäki, H., & Kotkas, T. (2006). Self-, peer and teacher-assessment of student essays. *Active Learning in Higher Education*, 7 (1), 51-62.
- López-Pastor, V. M.; Fernández-Balboa, J. M.; Santos, M. L., & Fraile, A. (2011). Students' self-grading, professor's grading and negotiated final grading at three university programmes: analysis of reliability and grade difference ranges and tendencies. *Assessment & Evaluation in Higher Education*, published online ahead of print publication, doi: 10.1080/02602938.2010.545868.
- Lu, J., & Law, N. (2011). Online peer assessment: effects of cognitive and affective feedback. *Instructional Science*, published online ahead of print publication, doi: 10.1007/s11251-011-9177-2.
- Magin, D., & Helmore, Ph. (2001). Peer and teacher assessment of oral presentation skills: How reliable are they? *Studies in Higher Education*, 26 (3), 287-298.
- Malini, Y., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35 (4), 435-448.

- Moskal, B. (2000). Scoring rubrics: what, when and how? *Practical Assessment, Research & Evaluation*, 7(3). Recuperado el 11 de mayo de 2009, <http://pareonline.net/getvn.asp?v=7&n=3>.
- Moskal, B. (2003). Recommendations for developing classroom performance assessments and scoring rubrics. *Practical Assessment, Research & Evaluation*, 8 (14). Recuperado el 11 de mayo de 2009, <http://pareonline.net/getvn.asp?v=8&n=14>.
- Moskal, B., & Leydens, J. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation*, 7(10). Recuperado el 11 de mayo de 2009, <http://pareonline.net/getvn.asp?v=7&n=10>.
- Norcini, J. J. (2003). Peer assessment of competence. *Medical Education*, 37(6), 539-543.
- Popham, J. W. (1997). What's wrong-and what's right-with rubrics. *Educational Leadership*, 55(2), 72-75.
- Roblyer, M. D., & Wiencke, W.R. (2003). Design and use of a rubric to assess and encourage interactive qualities in distance courses. *The American journal of distance education*, 17 (2), 77-98.
- Rohrbach, S. (2008). *Educational assessment in emerging areas of desing*. Design Research Society Biennial Conference 2008, Sheffield, UK.
- Sadler, P. M., & Good, E. (2006). The Impact of Self- and Peer-Grading on Student Learning. *Educational Assessment*, 11(1), 1-31.
- Sander, P. (2005). Researching our students for more effective university teaching. *Electronic Journal of Research in Educational Psychology*, 3 (1), 113-130.
- Schafer, W. D.; Swanson, G.; Bené, N., & Newberry, G. (2001). Effects of teacher knowledge of rubrics on student achievement in four content areas. *Applied Measurement in Education*, 14 (2), 151-170.
- Schamberr, J. F., & Mahoney, S. L. (2006). Assessing and improving the quality of group critical thinking exhibited in the final projects collaborative learning groups. *Journal of General Education*, 55 (2), 103-137.
- Stellmack, M. A.; Konheim-Kalskstein, Y.; Manor, J.; Massey, A. R., & Schmitz, J. A. (2009). An assessment of reliability and validity of a rubric for grading APA-style introductions. *Teaching of Psychology*, 36 (2), 102-107.
- Stix, A. (1997). *Creating rubrics through negotiable contracting and assessment*. US Department of Education. Recuperado el 11 de mayo de 2009, <http://www.interactiveclassroom.com/006%20-%20Creating%20Rubrics.pdf>.

- Struyven, K.; Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment and Evaluation in Higher Education*, 30 (4), 325-341.
- Struyven, K.; Dochy, F., & Janssens, S. (2008). The effects of hands-on experience on students' preferences for assessment methods. *Journal of Teacher Education*, 59 (1), 69-88.
- Taggart, G.; Phifer, S.; Nixon, J., & Wood, L. (2001). *Rubrics: A handbook for construction and use*. Lanham, MD: Scarecrow Press.
- Thaler, N.; Kazemi, E., & Huscher, C. (2009). Developing a rubric to assess student learning outcomes using a class assignment. *Teaching of Psychology*, 36 (2), 113-116.
- Tierney, R., & Marielle S. (2004). What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research & Evaluation*, 9 (2). Recuperado el 11 de mayo de 2009, <http://pareonline.net/getvn.asp?v=9&n=2>
- Tinsley, H.E., & Weiss, D. J. (2000). Interrater reliability an agreement. In H.E. Tinsley y S.D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling*, 95-124. San Diego: Academic.
- Zimmerman, B. J., & Schunk, D. H. (Eds.) (2001). *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd Ed.) Mahwah, NJ:Lawrence Erlbaum.

Apéndice 1. Rúbrica para la evaluación de las presentaciones orales con apoyo visual

1.- Principios básicos del modelo instruccional (30%)

CRITERIO	Muy adecuado (3)	Adecuado (2)	Básico/A mejorar (1)	Inadecuado (0)
Contenido modelo: Dominio y comprensión	Manifiesta una comprensión completa y profunda del tema	Muestra una buena comprensión del tema	Muestra una buena comprensión de partes del tema, pero no de alguna de ellas	No domina ni comprende el tema
Contenido modelo: Planificación y organización contenido	Discurso bien planificado, enlaza los tópicos de manera lógica y coherente	Sigue un esquema general adecuado, aunque alguna vez se pierde o necesita prestar más atención a detalles	Discurso difícil de seguir y conexiones incoherentes, lo que indica planificación pobre y que ha dedicado poco esfuerzo	No ha planificado
Contenido modelo: Integración elementos importantes	Incorpora de forma eficaz los elementos importantes del tema en toda la presentación	Incorpora de forma eficaz los elementos importantes del tema la mayor parte de la presentación	Escasa integración de los elementos fundamentales de la temática	Inadecuada integración

2.- Aplicación y ejemplificación del modelo instruccional (30%)

CRITERIO	Muy adecuado (3)	Adecuado (2)	Básico/A mejorar (1)	Inadecuado (0)
Aplicación práctica: Coherencia con modelo teórico	El ejemplo desarrollado es coherente con el modelo teórico	El ejemplo desarrollado es coherente con el modelo teórico, aunque necesitaría afinar algún aspecto concreto	El ejemplo desarrollado es coherente con el modelo, aunque diversos aspectos no quedan recogidos adecuadamente	El ejemplo desarrollado no es coherente con el modelo
Aplicación práctica: Clarificación modelo y motivación audiencia	El ejemplo resulta clarificador y motivante para la audiencia	El ejemplo no clarifica adecuadamente o resulta motivante para la audiencia	El ejemplo no clarifica adecuadamente y no resulta motivante para la audiencia	El ejemplo seleccionado es confuso y no se ajusta al modelo

3.- Apoyo visual (20%)

CRITERIO	Muy adecuado (3)	Adecuado (2)	Básico/A mejorar (1)	Inadecuado (0)
Apoyo visual: Número y explicación de las diapositivas	Cantidad de diapositivas adecuada y explicación pertinente de las mismas	Cantidad de diapositivas adecuada, aunque imprecisa explicación de alguna de ellas	Utiliza un número inadecuado de diapositivas, aunque intenta explicarlas	No cumple requisitos de número y adecuación en su explicación.
Apoyo visual: Adecuación de diapositivas	Todas tiene relación con el tema y conexión con el discurso, facilitando la comprensión del material y resultan amenas/captan la atención	Todas tiene relación con el tema y facilitan que la comprensión material, aunque algunas no están conectadas con el discurso o adolecen de amenidad	Aunque la mayoría de ellas están relacionadas con el tema, no favorecen su comprensión ni la motivación de la audiencia	Incorpora diapositivas no pertinentes con la temática, no facilitan la comprensión ni la motivación de la audiencia.
Apoyo visual: Legibilidad	No más de 5 palabras por línea, ni más de 5 líneas por diapositiva, letra legible para toda la audiencia, contraste	Letra legible para toda la audiencia, más de 5 líneas por diapositiva. Contraste adecuado de colores	Letra legible por el 85% audiencia aproximad., extensión excesiva en gran parte diapositivas. Inadecuado	Ilegible, uso de colores abrumador, abuso de texto y de párrafos

	adecuado de colores		contraste de colores	
<i>Apoyo visual: Relevancia y adecuación imágenes/ grafos/esquemas</i>	Relevantes, acordes con contenidos y con adecuada resolución	Relevantes, acordes con contenidos en la mayoría de ocasiones, pero poca resolución	Acorde con contenido, aunque no relevantes y de poca resolución	No utiliza o distraen
<i>Apoyo visual: Secuenciación, animaciones/transiciones diapositivas</i>	Animación puntos importantes, ayuda a centrar la atención en lo importante y evita distraerse	Orden lógico, animación de puntos no importantes o transiciones que dificultan atención	Orden lógico, aunque falta animación y sólo aplica transiciones en alguna diapositiva, o efectúa ambas de forma inadecuada	Ningún orden, excesivo o ningún uso animación y transición que facilita distracción

4.- Habilidades de comunicación en la presentación (20%)

<i>CRITERIO</i>	<i>Muy adecuado (3)</i>	<i>Adecuado (2)</i>	<i>Básico/A mejorar (1)</i>	<i>Inadecuado (0)</i>
<i>Habilidades verbales comunicación: Volumen y tono voz</i>	Suficientemente alto y con tono de voz adecuado para ser escuchado por toda la audiencia	Suficientemente alto para ser escuchado por todos y tono de voz adecuado alrededor del 90%	Suficientemente alto para ser escuchado alrededor del % del tiempo por todos y/o tono inadecuado (monótono, aburrido)	Demasiado bajo ser escuchado por todos y excesivamente monótono
<i>Habilidades verbales de comunicación: Claridad en el habla</i>	Habla claro la práctica totalidad del tiempo.	Habla claro el 90-95% del tiempo, aunque se equivoca al pronunciar algunas palabras	Habla claro 85-90% del tiempo y pronuncia mal numerosas palabras	Poca claridad. A menudo masculla y pronuncia mal palabras
<i>Habilidades verbales de comunicación: Pausas y uso de "coletillas"</i>	Usa las pausas correctamente y al final de las oraciones. Utiliza frases completas en más del 95% ocasiones	Usa las pausas correctamente, aunque introduce algunas de forma inesperada/ inadecuada, utiliza en diversas ocasiones coletillas (um..., eh...etc). Utiliza frases incompletas en más del 5% ocasiones	Número de pausas inadecuado –por exceso o defecto–, introduce algunas en momentos inesperados o utiliza frecuentemente coletillas (um..., eh...etc). Utiliza muchas frases incompletas	No usa pausas de manera intencionada. Abuso de coletillas. Gran cantidad de frases incompletas
<i>Conducta no verbal: Postura y contacto visual</i>	Espalda recta, postura relajada y confiada, establece contacto visual con todos	Postura erguida y establece contacto visual con todos, aunque se observan algunos indicadores de tensión/desinterés	A veces mantiene postura erguida y mantiene contacto visual con la audiencia, pero menos del 90% del tiempo	No mantiene contacto visual y la postura corporal es inadecuada