



UNA APROXIMACIÓN ANALÍTICA DE LA INVESTIGACIÓN SOBRE LA EVALUACIÓN DE LAS RÚBRICAS. POSIBILIDADES, LIMITACIONES Y RETOS PARA SU USO

MÓNICA GARCÍA HERNÁNDEZ
UNIVERSIDAD PEDAGÓGICA NACIONAL
monigarher@gmail.com

MARÍA DEL CARMEN VELEROS VALVERDE
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
cvelerosv@gmail.com

RESUMEN

En el último cuatrienio se ha desarrollado de manera sobresaliente la investigación sobre las rúbricas, con aportaciones interesantes de trabajos iberoamericanos y nacionales. En México ha crecido la atención por su uso no sólo en el contexto del aula, sino también de las evaluaciones a gran escala. La evaluación de la validez y confiabilidad de los resultados obtenidos de una rúbrica, es un punto de debate importante en la literatura especializada que no puede soslayarse, dados los retos e implicaciones sociales, metodológicas, de política de la evaluación y éticos de la misma. Con este propósito se plantea en el presente escrito, como parte de un proyecto más amplio de investigación sobre rúbricas, el análisis de su evaluación válida y confiable, que destaque logros, dificultades y posibilidades; donde la validez y confiabilidad son visualizadas desde varias dimensiones. Destacan la diversidad de posturas metodológicas, el todavía incipiente pero revelador desarrollo de acercamientos mixtos, validez sustantiva y consecuencial de los trabajos revisados. Se concluye sobre la necesidad de estimar tales conceptos de manera múltiple acorde al contexto y fines de la evaluación en que se pretenda usar rúbricas.

Palabras clave: Investigación, rúbricas, validez, confiabilidad, evaluación auténtica

POR QUÉ LA INVESTIGACIÓN SOBRE RÚBRICAS

En el ámbito académico, la investigación sobre rúbricas es un campo de estudio relativamente reciente que surge hace unos quince años en países de habla inglesa, con una historia aún más corta entre los





estudiosos hispanos y latinoamericanos. Se cuenta en la literatura especializada en el tema con tres estados de conocimiento (Johnson y Svingy, 2007, Reddy y Andrade, 2010, Panadero y Johnson, 2013) que incluyen esencialmente investigaciones anglosajonas. Más de las tres cuartas parte de la investigación sobre rúbricas se ha producido a partir del 2011 y más de la mitad en los dos últimos años con un aumento importante de aportaciones iberoamericanas (García y Veleros, 2014).

Actualmente en México se ha comenzado a utilizar rúbricas como parte de acciones de política educativa en el contexto del reconocimiento a los discursos de corte constructivista y competencial respecto al aprendizaje, la práctica docente y la evaluación educativa. Buen ejemplo de ello son la generación de documentos por parte de la Secretaría de Educación Pública para la formación docente (Dirección General de Educación Básica, 2013) y el proceso de ingreso a nivel medio superior (Secretaría de Educación Pública, 2014); así como iniciativas de organismos de evaluación como el Centro Nacional de Evaluación para la Educación Superior (CENEVAL) en la parte del examen de Competencia Comunicativa y Pensamiento Crítico de Nivel Licenciatura y del Instituto Nacional para la Evaluación de la Educación (INEE) con rúbricas para evaluar el grado de congruencia de la práctica de educadoras (Zúñiga, Vilchis, Álvarez, López y Aguilera, 2013).

Consideramos necesario el análisis sobre el uso de las rúbricas, dadas sus implicaciones para una política de evaluación y para las prácticas evaluativas del aprendizaje en el aula. Con el propósito de contribuir al estudio sobre este uso, se llevó a cabo un proyecto de investigación sobre la investigación de las rúbricas¹, el cual comprendió tres etapas: elaboración de una base datos electrónica con 59 fichas de análisis de literatura especializada seleccionada del 2005 al 2014 como resultado de un proceso de revisión de 142 documentos y debate colectivo a lo largo de 18 meses²; análisis cuantitativo y cualitativo de esta literatura elegida complementado con la entrevista de cuatro expertos nacionales con publicaciones en el tema; finalmente el desarrollo de un programa de intervención con 30 docentes de educación básica, media superior y superior para el diseño de rúbricas y el seguimiento de sus prácticas de uso en el salón de clase.

La base de datos y el análisis de la literatura finalmente depurada se organizó a partir de seis criterios de análisis: año de publicación, idioma y lugar de investigación, metodología seguida, sujetos diseñadores de rúbricas, nivel escolar y carácter del estudio (dividido a su vez en aportaciones conceptuales y trabajos aplicados). Dada la amplitud del mencionado proyecto y por motivos de





espacio, en este escrito se presenta un avance sobre el análisis de uno de los aspectos relevantes en el desarrollo de rúbricas: la evaluación de su validez y confiabilidad.

LA IMPORTANCIA DE LA EVALUACIÓN DE RÚBRICAS

La propuesta de las rúbricas, matrices o plantillas de valoración surge con el movimiento de evaluación auténtica, como parte de los enfoques experienciales y constructivistas del aprendizaje, especialmente los de carácter sociocultural y situado; desde los cuales se destacan una visión holística, multidimensional, procesual y compartida de la evaluación de tareas generativas para la calificación y desarrollo contextualizados de desempeños complejos. Con base en estos marcos conceptuales, particularmente pertinentes son las definiciones de rúbricas como guías de trabajo, estrategias de evaluación (Díaz Barriga, 2006), una forma de andamiaje (*hard scaffolding* según Giacumo, Saveyne y Smith, 2013), un enfoque de autoevaluación (Panadero y Romero, 2014) y una herramienta de integración del trinomio enseñanza, aprendizaje y evaluación (De la Cruz, 2011).

En congruencia con los anteriores planteamientos y de acuerdo a Picón (2013) concebimos a la rúbrica como un instrumento compuesto por una tarea evaluativa auténtica y su correspondiente escala de valoración, la cual cobra sentido en cuanto instrumento de interpretación del desempeño del estudiante en la realización de dicha tarea.

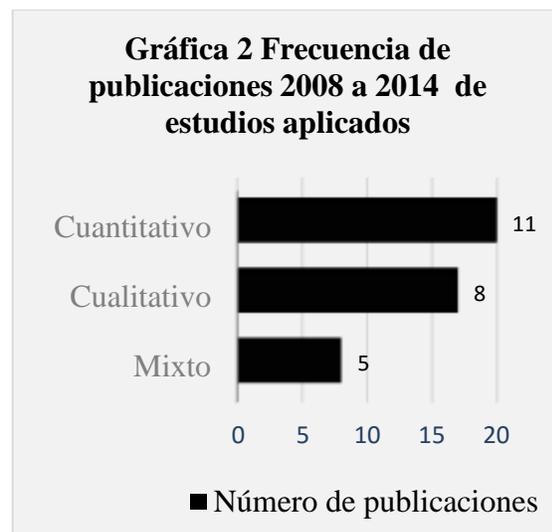
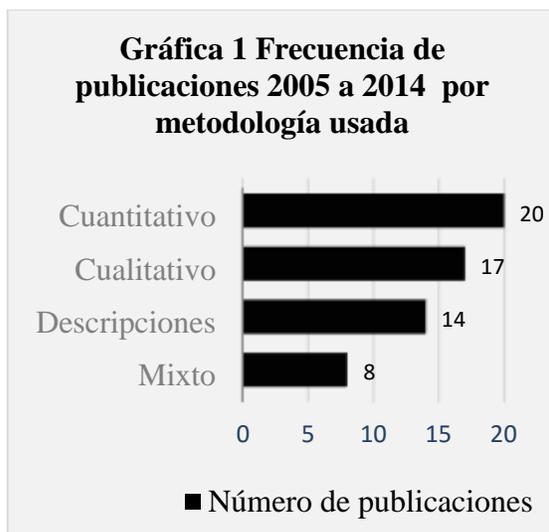
Como todo instrumento educativo, lo que otorga carácter científico a la información derivada de una rúbrica es que se demuestre que ésta evalúa lo que pretende evaluar (validez) y que lo hace de manera precisa o con el mínimo de error o sesgo (confiabilidad). Hay distintas posturas para el logro de estas cualidades, cuyos criterios se derivan y varían de acuerdo a los fines y contextos con que se realice una evaluación: por una parte existen las evaluaciones de gran escala motivadas por las necesidades de información que los responsables de las políticas tienen para la organización de los sistemas educativos, por otra parte contamos con las evaluaciones de aula vinculadas a decisiones específicas de docentes respecto a las necesidades de los estudiantes en un determinado currículo (Shepard, 2006). En el primer caso, la rúbrica es empleada como un instrumento de carácter sumativo, donde su validación y confiabilidad se estiman por medio de análisis estadísticos, además de que se busca la generalización de los resultados independientemente de las condiciones. En cambio en el segundo, ésta constituye una herramienta de tipo formativo, que emplea datos cualitativos, cuyo uso es dependiente del contexto.





Existe tensión y se mantiene la discusión sobre si la estandarización de las rúbricas, es la vía legítima y única de rigor metodológico. Por una parte se plantea que tal estandarización implica una distorsión en los fines y principios constructivistas y de evaluación auténtica de rúbricas, debido al papel de selección de los evaluados que deviene en la exclusión del acceso a la educación de determinados grupos sociales. Por otra parte, se asiste a la institucionalización del manejo de las rúbricas en este tipo de evaluaciones, con retos de carácter metodológico e implicaciones en los resultados por su impacto social y de política educativa en los sistemas de acreditación y certificación de los sistemas escolares.

La propuesta mixta o combinación de métodos cuantitativos y cualitativos para el desarrollo de la validez y confiabilidad, constituye otra posibilidad para la evaluación de las rúbricas. De acuerdo a los criterios de metodología seguida y carácter del estudio de la muestra de los 59 trabajos seleccionados, la investigación sobre rúbricas se caracteriza por la diversidad de enfoques metodológicos con que se hace. Como se ilustra en la gráfica 1 es ligeramente mayor el número las investigaciones de corte cuantitativo (33.89%) sobre el cualitativo (28.81%); le siguen los documentos de exposiciones y descripciones (23.72 %) y de forma todavía incipiente, trabajos (13.55%) con una metodología mixta que combina el acercamiento cualitativo y el cuantitativo. Con base al criterio de estudios de carácter aplicado y del año de su publicación (2008 para investigaciones nacionales y 2010 desde el primer estado de conocimiento sobre rúbricas en contexto anglosajón), se depuró una muestra de 24 investigaciones (ver gráfica 2) para el análisis de la validez y confiabilidad de los estudios en rúbricas.





Como parte del debate de la evaluación al servicio de la mejora educativa, mismo que aplica para las rúbricas, se plantea que el docente priorice la evaluación del aprendizaje en el contexto del aula y que los resultados de las evaluaciones a gran escala sean complemento para el desarrollo de su trabajo y toma de decisiones (Martínez, 2008 citado en Shepard, 2006).

INVESTIGACIÓN SOBRE LA VALIDEZ DE LAS RÚBRICAS

Si bien las investigaciones sobre rúbricas presentan problemas metodológicos, debido a la poca claridad y ausencia de elementos básicos como tipo y calidad del diseño, características de la muestra, especificación de estándares de evaluación, del procedimiento, y reporte de resultados (Panadero y Romero, 2014), también es importante considerar que los conceptos, criterios y estrategias respecto a la validez y confiabilidad de los resultados de un instrumento se han diversificado, lo que se aplica al caso de las rúbricas y es congruente con la variedad de metodologías sobre la investigación de las mismas. En este tenor de ideas, con base en la metodología del análisis narrativo del contenido³ reportada por Panadero y Tapia (2013) y la adaptación de los criterios de Mesick⁴ (1996 citados por Johnson y Svingy, 2007), realizamos el análisis sobre la validez de los resultados de las 24 investigaciones seleccionadas sobre trabajos aplicados.

A partir de los aspectos de validez seguidos (ver tabla 1), se confirma la tendencia del aspecto del contenido en más de la mitad de las investigaciones. En menor medida, siguen los aspectos de generalización y consecuencias. Es escaso el interés de los aspectos de validez estructural y sustantivo (ver tabla 2). Tabla 1 Descripción de aspectos de validez para análisis de investigaciones sobre rúbricas

Aspectos de validez	Descripción
Contenido	Relevancia y representatividad de conocimientos y habilidades del área de dominio de una rúbrica
Generalización	Extensión con que se extienden las interpretaciones la rúbrica a grupos, ocasiones, tareas, entre otros.
Estructural	Consistencia entre la teoría y las dimensiones, criterios y puntaje de la rúbrica.
Sustantivo	Consistencia entre racionalidad teórica, evidencia empírica y resultados del proceso de uso.
Consecuencial	Implicaciones y evidencia de impacto positivo el desarrollo de rúbricas





Tabla 2 Frecuencia de investigaciones por aspectos de validez y tipo de metodología

Aspectos de validez	Metodología			Número de estudios
	Cuantitativa	Cualitativa	Mixta	
Contenido	6	4	3	13
Generalización	2	1	1	4
Estructural	1	0	0	1
Sustantivo	1	1	0	2
Consecuencial	1	2	1	4

INVESTIGACIÓN SOBRE LA CONFIABILIDAD DE LAS RÚBRICAS

A pesar de la falta de acuerdo sobre si la confiabilidad intra sustentada en que los propios investigadores son los evaluadores, es una limitación metodológica que lleva a sesgos o una condición pertinente, es la confiabilidad interjueces la que ha recibido más atención.

En las estrategias de confiabilidad inter también se confirma la tendencia del aspecto del consenso como el de mayor uso (ver tabla 3), ya sea a partir de expertos y los propios participantes evaluados en el estudio. Por lo que respecta a la evidencia de las calificaciones, los estudios mixtos llevan a cabo tanto mediciones estadísticas como cualitativas en equilibrio con el aspecto de consenso, los estudios cuantitativos continúan centrándose en medidas estadísticas sobre las de otro tipo y los estudios cualitativos son los que menos precisan evidencias de sus calificaciones.

Tabla 3 Frecuencia de investigaciones por aspectos de confiabilidad y tipo de metodología

Aspectos	Metodología			Número de estudios
	Cuantitativa	Cualitativa	Mixta	
Un evaluador	0	1	0	1
Consenso	6	5	4	15
Medición	5	3	4	11





REFLEXIONES FINALES: USO DE RÚBRICAS

Consideramos que a pesar de la necesidad de desarrollar un mayor rigor metodológico en las investigaciones sobre rúbricas, es indispensable distinguir las diferencias de estimación de la validez y confiabilidad de las evaluaciones en el aula respecto a las evaluaciones a gran escala, así como las condiciones exógenas y no sólo endógenas (Díaz Barriga y De la Cruz, 2011) en el uso válido y confiable de las rúbricas.

La mayoría de usos de las rúbricas constituye evaluaciones en el aula con fines formativos, realizadas de manera constante y dinámica, donde lo más importante es la suficiencia de la información sobre el desempeño para el andamiaje o apoyo del proceso constructivo del estudiante, sin necesidad de que el docente tenga que fundamentarse en calificaciones estables e interpretaciones fijas sobre el aprendizaje. En este sentido, la validez que interesa es la pertinencia del contenido con la contextualización del currículo, que conduzca o tenga por consecuencia la toma pertinente de decisiones docentes para y por el avance del aprendizaje del estudiante. Lo anterior implica una visión sistémica de estrategias de evaluación (Medida, 2013) y su articulación con la enseñanza. Visión que sería pertinente incluir en las evaluaciones a gran escala que usen rúbricas.

La evaluación sobre la validez y confiabilidad de las rúbricas como herramientas de enseñanza y evaluación, ha de visualizarse no sólo de manera situada en las condiciones y fines, sino también respecto al ejercicio ético regulado de su uso, sin el cual se corre el riesgo de favorecer prácticas simuladoras y encubiertas de la evaluación con consecuencias negativas para el aprendizaje. Para ello planteamos que los criterios de equidad como respecto a la diversidad, transparencia como conocimiento preciso y compartido de las calificaciones y democracia como consenso responsable y múltiple de evidencias (Picón, 2009) formen parte de una evaluación válida y confiable con rúbricas en el trabajo docente en el aula, además de no dejarse a un lado en las evaluaciones de alta escala, aunque esto último sea hasta ahora más un ideal que una realidad.

NOTAS

¹ Proyecto "El uso de las rúbricas en educación superior: conceptos y aplicaciones" con registro 011/II/2012, Área 3 Aprendizaje y Enseñanza en Ciencias, Humanidades y Artes, Universidad Pedagógica Nacional, Unidad Ajusco.





² Se siguieron las siguientes etapas: (1) recopilación de literatura generada de 1998 a 2012 en tres bases de datos, en tres congresos especializados, capítulos de libro y dos bancos de tesis; (2) selección de literatura a partir primero de una revisión general y una segunda revisión detallada con la ayuda de una ficha analítica, organizada por datos de identificación, propósito, conceptos (clave), metodología, fuentes (principales), aplicaciones, observaciones y resumen; (3) integración de la información para primer borrador sobre la investigación de rúbricas; (4) recopilación de la literatura publicada generada en 2013 y 214, que abarcó seis bases de datos más; (5) selección y análisis de esta literatura a partir de la ficha analítica diseñada; (6) elaboración de base analítica electrónica de datos y (7) versión final de la revisión analítica sobre la literatura en investigación de rúbricas.

³ Consistió en la elaboración de tablas analíticas comparativas que comprendieron diseño, muestra, tarea, contenido rúbrica y tipo de ayuda, procedimientos, resultados, variables, efectos y problemas metodológicos.

⁴ De los criterios de Mesick se eliminó el aspecto externo, dada la ausencia de investigaciones que incluiría junto con la rúbrica otros instrumentos de evaluación.

Recuerde que las notas son contabilizadas para la extensión máxima de 3000 palabras para las ponencias.

REFERENCIAS

De la Cruz, F.G. (noviembre, 2011). La construcción y aplicación de rúbricas: una experiencia en la formación de psicólogos educativos. *Observar* 5, 21-11. Recuperado de <http://www.odas.es/site/new.php?nid=37>

Dirección general de Educación Básica (2013). Las estrategias y los instrumentos de evaluación desde el enfoque formativo 4. México: SEP. Recuperado de <http://basica.sep.gob.mx/C4%20HERRAMIENTAS-ESTRATEGIAS-WEB.pdf>

Díaz Barriga, F. (2006). *Enseñanza Situada. Vínculo entre la escuela y la vida*. México: McGrawHill.

Díaz Barriga, F. y De la Cruz, F.G. (2011). Rúbricas en la evaluación de competencias y aprendizaje completo. Alcances y restricciones en educación superior. En Bujan, V.K., Rekalde, R. I. y Aramendi, J.P. (Edits.). *La evaluación de competencias en la educación superior. Las rúbricas como instrumentos de evaluación* (13-36). Bogotá: Ediciones de la U.

García, H.M. y Veleros, V. M. C. (2014). La investigación sobre las rúbricas. Revisión general de literatura. *Memorias Congreso Internacional de Educación. Evaluación*. Tlaxcala: Universidad Autónoma de Tlaxcala.





- Giacumo, A.L., Saveyne, W. y Smith, N. (2013). Facilitation prompts and rubrics on higher-order thinking skill performance found in undergraduate asynchronous discussion boards. *British Journal of Educational Technology*, 44(85), 774-794. Recuperado de <http://www.ebscohost.com>
- Johnson, A. y Svingy, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130-144. Recuperado de <http://www.ebscohost.com>
- Medina, L. (2013). La evaluación en el aula: reflexiones sobre sus propósitos, validez y confiabilidad. *Revista Electrónica de Investigación Educativa*, 15(2), 34-50. Recuperado de <http://redie.uabc.mx/vol15no2/contenido-medina.html>
- Panadero, E. y Jonson, J A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129-144. Recuperado de <http://www.ebscohost.com>
- Panadero, E. y Romero, M. (2014). To rubric or not to rubric. The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice*, 21(2), 133-148. Recuperado de <http://www.ebscohost.com>
- Panadero, C.E. y Tapia, A.J. (2013). Revisión sobre autoevaluación educativa: evidencia empírica de su implementación a través de la autocalificación sin criterios de evaluación, rúbricas y guiones. *Revista de Investigación en Educación*, 11 (2), 2013, 172-197.
- Picón, J.E. (septiembre-diciembre 2013). La rúbrica y la justicia en la evaluación. *Íkala Revista de lenguaje y cultura*, 79-94. Recuperado de <http://aprendeenlinea.udea.edu.co/revistas/index.php/ikala/article/viewFile/14080/15851>
- Reddy, M. y Andrade, H. A (Julio, 2010). Review of rubric use in higher education. *Assessment & evaluation in higher education*, 35, (4), 435-448. Recuperado de <http://www.ebscohost.com>
- Secretaría de Educación Pública (2014). *Rúbrica de Ingreso al Servicio Profesional Docente para la Educación Media Superior de Planeación Didáctica y EXPRESE*. Guía del sustentante. México: SEP.





Zúñiga, P. L. Y H., Vilchis, G.J.E., Álvarez, L. A.C., López, G. A.Y. y Aguilera, G.M.Z. A. (2013). Prácticas pedagógicas y desarrollo profesional docente en preescolar. Reporte técnico. México: Instituto Nacional para la Evaluación de la Educación.
<http://publicaciones.inee.edu.mx/buscadorPub/P1/D/240/P1D240.pdf>

