

LA RÚBRICA Y LA JUSTICIA EN LA EVALUACIÓN

[THE ROLE OF RUBRICS IN FAIR ASSESSMENT PRACTICES]

Édgar Picón Jácome

Grupo de Investigación Acción y
Evaluación en Lenguas Extranjeras GIAE
Escuela de Idiomas
Universidad de Antioquia UdeA
Calle 70 No. 52-21. Medellín, Colombia
Email: epjacome59@hotmail.com

RESUMEN

En este artículo se sustenta la tesis de que la utilización de rúbricas consensuadas garantiza en gran medida la promoción de prácticas evaluativas justas. Esta aseveración se basa en el argumento de que la creación y aplicación de rúbricas incrementa la *validez* y la *transparencia* en la evaluación, a la par que su *diseño* consensuado promueve la democracia, posibilita el impacto positivo en el aprendizaje o efecto *washback* y facilita prácticas *equitativas* —siendo todos estos principios inherentes a la justicia. Se enfatiza en un concepto de rúbrica que integra una tarea auténtica y su escala de valoración, como procedimiento de evaluación alternativa, dentro de un enfoque crítico.

Palabras clave: rúbricas, justicia, democracia, evaluación alternativa, evaluación formativa, lenguas extranjeras

ABSTRACT

In this article, I support the thesis that utilizing rubrics, discussed and agreed upon teachers and students, guarantees the promotion of fair assessment practices. This claim is supported by the argument that designing and applying scoring rubrics enhances *validity* and *transparency* in evaluation, while their consensual design supports *democracy*, promotes positive *washback* and facilitates *equitable* assessment practices—all of them principles inherent to fairness. A concept of rubric that integrates an authentic task and its scoring scale, applied as an alternative assessment procedure within a critical approach, is emphasized.

Keywords: rubrics, fairness, democracy, alternative assessment, formative assessment, foreign languages

Este artículo es producto de la investigación "Definición de criterios comunes para la evaluación de la competencia comunicativa en programas de inglés como lengua extranjera" inscrito en el Sistema Universitario de Investigación de la Universidad de Antioquia, y cofinanciado por la Escuela de Idiomas de la Universidad de Antioquia y la Escuela de Educación y Pedagogía de la Universidad Pontificia Bolivariana, Sede Medellín.

Introducción

La evaluación del aprendizaje en el campo de las lenguas extranjeras es un tema complejo para el que los profesores necesitan información y desarrollo profesional. En el contexto local colombiano, Frodden, Restrepo y Maturana (2004), y Arias y Maturana (2005), a partir de un estudio llevado a cabo en programas de enseñanza del inglés como lengua extranjera de dos universidades públicas, reportan una reiterada confusión de los docentes participantes respecto al constructo a evaluar y a las características y los propósitos de evaluación de las tareas lingüísticas. Por su parte López y Bernal (2009), en un estudio sobre las percepciones de los profesores de inglés colombianos respecto a la evaluación en el aula, de manera preocupante encontraron que la gran mayoría de las universidades participantes en su estudio (20 de 27) no ofrecían cursos en evaluación como parte de sus programas de pregrado (p. 62).

80

Esta condición parece afectar las actitudes de los docentes de lenguas respecto a la evaluación y afecta las decisiones que estos toman en el momento de promover a sus estudiantes a niveles superiores. De hecho, López y Bernal (2009) reportaron una diferencia sustancial de las percepciones entre los docentes con formación en evaluación y los que no la tenían: mientras que los primeros consideraban la evaluación como una herramienta útil que les permitía tener un impacto positivo en el proceso de enseñanza-aprendizaje, los últimos la consideraban simplemente una herramienta de poder y control y un proceso sumativo obligatorio con el objetivo de generar una nota (p. 62). Paralelamente, Arias y Maturana (2005) concluyeron que la falta de precisión en cuanto al constructo a evaluar podría afectar las decisiones de promoción de los profesores, las que en todo caso se tornaban cuestionables (p. 86).

Este artículo es un aporte al desarrollo profesional de los docentes de lenguas a partir de una reflexión en torno a aspectos relacionados

con la evaluación del aprendizaje en dicho campo. A través de esta reflexión busco motivar a los docentes a llevar a cabo una evaluación justa en el aula —lo que significa que sea equitativa, válida, transparente, con propósitos formativos y enmarcada en principios democráticos— y promover entre ellos el diseño y aplicación de rúbricas en sus clases. Me parece importante enfatizar en la necesidad de que los profesores se formen en el desarrollo de procedimientos evaluativos, y en el estudio de constructos tales como la *competencia comunicativa* (Instituto Cervantes, 2002; Savignon, 2001) y la *habilidad lingüística* (Bachman & Palmer, 1996), u otros, coherentes con los enfoques metodológicos que sustenten sus prácticas de enseñanza, para que sean ellos mismos los que diseñen e implementen formas válidas y confiables de evaluación de manera informada y autónoma. En ese sentido este artículo evidencia lo que se puede lograr a través de procesos de formación profesional docente continuos como los realizados por medio de grupos de estudio de profesores (Birchak, Connor, Crawford, Kahn, Kaser, Turner & Short, 1998, 1998; Frodden & Picón, 2005).

La rúbrica

El término *rúbrica* se origina como una traducción de la palabra inglesa *rubric*. En el campo de las pruebas evaluativas tradicionales, denotadas por el término *testing*, Bachman y Palmer (1996) presentan la rúbrica como el conjunto de características de una prueba que estructuran las tareas evaluativas que la componen. Estos factores incluyen la estructura de la prueba en sí; las instrucciones; el tiempo determinado, tanto para la realización de la prueba en general, como de cada tarea; y el método de evaluación y calificación (Bachman & Palmer, 1996, p. 50). Esta definición amplia de rúbrica aplica para pruebas en las que se mezclan ítems de respuesta cerrada —del tipo de escogencia múltiple, por ejemplo— y tareas de respuesta abierta.

Por su parte, Mansoor y Grant (2002) definen la rúbrica como “un instrumento [de evaluación] que especifica el desempeño general esperado y los diversos niveles de competencia a los que los aprendices pueden llegar en el desarrollo de una habilidad dada” (p. 33).¹ Esta segunda definición está de acuerdo con lo que proponen otros autores que se refieren a la rúbrica como un instrumento de evaluación pertinente para tareas evaluativas auténticas —tales como entrevistas, ensayos, etc.— y en general dentro de propuestas de evaluación más alternativas (Goodrich-Andrade, n.d.; Mertler, 2001; Moskal, 2000; O’Malley & Valdez, 1996; Panadero & Jonsson, 2013; Reddy & Andrade, 2010; Stevens & Levi, 2005). Desde este mismo enfoque, autores como Brown (2004) y Genesee y Upshur (1996), aunque no utilizan el término rúbrica, proponen métodos de evaluación —*scoring methods*— que incluyen escalas de valoración² —*scoring scales*— analíticas y holísticas para evaluar el mismo tipo de tareas.

En este artículo me refiero a la rúbrica en los términos de Mansoor y Grant (2002), agregando que debe además ir acompañada de información específica que incluya todos los detalles que los estudiantes necesitan saber para completar una tarea de manera exitosa. Es así que la tarea evaluativa forma parte indispensable de la rúbrica ya que la escala de valoración cobra sentido en cuanto instrumento de interpretación del desempeño del estudiante en la realización de dicha tarea. El término rúbrica en esta reflexión se refiere entonces a un instrumento de evaluación compuesto por una tarea evaluativa auténtica (O’Malley & Valdez, 1996) y su correspondiente escala de valoración, la cual estaría acorde con el tipo de rúbrica que Moskal (2000) denomina “*task specific*”.

La justicia en la evaluación

El principio de justicia ha sido tradicionalmente aplicado en la evaluación desde una perspectiva positivista en la que los evaluadores se limitan a garantizar a los evaluados *iguales condiciones* en la realización de una prueba. Dicha *igualdad* se refiere a aspectos como el tiempo, el espacio y los materiales. Puesto que los estudiantes son individuos con historias de vida y estilos de aprendizaje diversos, que en muchas ocasiones pertenecen a culturas completamente diferentes, y que no han tenido necesariamente las mismas oportunidades de acceso a una educación de calidad, no es suficiente con garantizar igualdad en dichos términos para ser justos.

Suskie (2002) aboga, en cambio, por la promoción de la *equidad* en la evaluación para garantizar la justicia. Parafraseando a Lam, Suskie define una evaluación justa como “aquella en la que se les da a los estudiantes *oportunidades equitativas para demostrar lo que saben*”³ (Lam, 1995, citado en Suskie, 2002, p. 5), lo cual significa que deberíamos evaluarlos utilizando los métodos y procedimientos más apropiados de acuerdo con condiciones tales como su “conocimiento previo, experiencia cultural y estilo cognitivo” (p. 5). Por su parte, Shohamy (2001) promueve un compromiso mayor, por parte de los implicados, en la lucha por garantizar prácticas evaluativas justas que la autora encuentra inmersas dentro de lo que llama *evaluación democrática como alternativa* (p. 376).

Dentro del marco de esta reflexión, y siguiendo esencialmente las ideas de Suskie (2002) y Shohamy (2001), defino la justicia como un principio que implica el desarrollo de prácticas evaluativas que garanticen la equidad y que

¹ Todas las traducciones presentadas en este artículo son autoría del escritor.

² El término *escala de valoración* se refiere a lo que otros autores llaman en español *baremo* o *parrilla*. A lo largo del artículo se utilizan los términos *rúbrica* —en las citas y referencias en este campo— y *escala de valoración*, para ser más específico al referirse al concepto de rúbrica definido en este apartado.

³ Énfasis del autor.

sean además válidas y transparentes. Considero asimismo que la búsqueda de la justicia debe incluir el desarrollo de formas de evaluación que sigan principios democráticos, es decir participativos e incluyentes, en la toma de decisiones dentro del aula. Estos principios y cualidades se tornarían a su vez en lineamientos a seguir en el diseño de los procedimientos de un sistema de evaluación que garantice un impacto positivo en los procesos de enseñanza y aprendizaje. Dicho impacto, que se traduce en una evaluación para el aprendizaje, debe ser el principal objetivo que guíe todo evento evaluativo dentro del aula.

La rúbrica y la justicia

A continuación defino los términos claves que sustentan el concepto de justicia que acabo de presentar: la equidad, la validez, la transparencia, el impacto y la democracia. Presentando paralelamente los argumentos que sustentan mi tesis de que el diseño y aplicación de rúbricas consensuadas con los alumnos incrementa el principio de justicia en las prácticas evaluativas, me propongo guiar al lector en el entendimiento de dichos conceptos.

La rúbrica y la equidad.

De acuerdo con Suskie (2002), la justicia está en gran medida enmarcada por la búsqueda de la equidad en la evaluación. Según el Instituto de Estudios sobre Desarrollo y Cooperación Internacional, HEGOA (n.d.),

La equidad introduce un principio ético o de justicia en la igualdad ... nos obliga a plantearnos los objetivos que debemos conseguir para avanzar hacia una sociedad más justa. Una sociedad que aplique la igualdad de manera absoluta será una sociedad injusta, ya que no tiene en cuenta las diferencias existentes entre personas y grupos.

De manera consecuente, Suskie (2002) plantea que ser justos implica proporcionar a los estudiantes no sólo muchas oportunidades, sino también muchas formas diferentes de evaluación, con el fin de garantizarles diversas posibilidades

de demostrar lo aprendido. Ser equitativos implicaría entonces tener en cuenta las diferencias de nuestros evaluados en el momento de diseñar los procedimientos a seguir. Estas diferencias pueden ser de tipo cultural o cognitivo —de hecho la literatura da cuenta de diversidad en los estilos y estrategias de aprendizaje de los alumnos además de identificar diferentes formas de enseñanza en diferentes contextos culturales (ver Oxford, 2001).

La equidad se puede lograr mediante la aplicación de una perspectiva múltiple en la evaluación. El principio de la multiplicidad, desarrollado por Shohamy (1998), se refiere a la necesidad de tomar decisiones informadas en el momento de diseñar los procedimientos en una evaluación (Shohamy, 1998; Shohamy & Imbar, 2006). El proceso de diseñar rúbricas, en los términos en que ellas se definen en esta reflexión, se ajusta a las fases seguidas en el diseño de procedimientos evaluativos desde dicha perspectiva. Dependiendo de la forma en que se utilice, la rúbrica permite lograr propósitos tanto formativos como sumativos en la evaluación y se ajusta a procedimientos evaluativos diversos que pueden ser tareas auténticas de producción oral —tales como entrevistas, juegos de rol y exposiciones— o de producción escrita —tales como composiciones cortas, diarios o ensayos—. Esta gama de posibilidades le permite al docente evaluar a estudiantes con diferentes características a través de diferentes procedimientos, y por tanto ser equitativo.

Rúbrica, transparencia y validez.

Suskie (2002) enfatiza en la necesidad de ligar la evaluación a la enseñanza para garantizar la justicia e insta a los profesores a tomar en cuenta factores como los estándares y los criterios establecidos para pasar al siguiente curso en el momento de calificar el trabajo de los estudiantes; con base en esta propuesta de la autora, en este apartado argumento que una evaluación justa es también una evaluación válida.

De acuerdo con Williams (2003) dicha conexión entre enseñanza y evaluación corresponde al principio de validez: “[La validez] se debe entender como la concordancia entre lo que se está enseñando y lo que se está midiendo” (p. 231). Por su parte, Brown (2004) asevera que “no hay una forma final y absoluta de medir la validez, pero se pueden esgrimir evidencias de diferentes tipos para argumentarla” (p. 22), y entre ellas menciona evidencias relacionadas con el constructo y con el contenido.

Según Brown “un constructo es cualquier teoría, hipótesis, o modelo que intenta explicar los fenómenos observados en nuestro universo de percepciones” (p. 25). Fulcher y Davidson (2007), en el campo de la enseñanza y aprendizaje de lenguas, utilizan el término *constructo* para referirse a la descripción de los componentes de un modelo. Sin embargo, advierten sobre la confusión en la literatura acerca de los términos *modelo* y *marco*, y proponen diferenciarlos, siendo los modelos las descripciones teóricas de lo que significa la habilidad de comunicarse en una segunda lengua, mientras que los marcos la selección, tomada de un modelo, de las habilidades a ser evaluadas en determinado contexto (p. 36). En otras palabras, el constructo se refiere al *qué* en la evaluación, y tiene una dimensión teórica —el modelo de lengua— y una práctica —el marco determinado para un momento evaluativo concreto—. Ejemplos de modelos serían la *competencia comunicativa* (Savignon, 2001) y la *habilidad lingüística* (Bachman & Palmer, 1996).⁴ Finalmente, Bachman y Palmer (1996) se refieren a la validez del constructo afirmando que “pertenece a lo significativo y apropiado de las interpretaciones que hacemos con base en las calificaciones” (p. 21).

En el campo de la enseñanza y aprendizaje de lenguas extranjeras la validez de constructo presenta probablemente el mayor reto a lograr en la evaluación debido a que es en este aspecto donde profesores y administradores parecen estar más confundidos (ver Arias & Maturana, 2005). Definir un constructo coherente demanda de los implicados en el proceso evaluativo claridad en cuanto al modelo de lengua que sustenta un programa. En este aspecto los administradores y docentes requieren actualizarse continuamente a través de programas de desarrollo profesional tales como los grupos de estudio (Birtchack et al., 1998; Frodden & Picón 2005).⁵

Respecto a la validez de contenido, Brown (2004) asevera:

Si una prueba es en realidad un ejemplo del asunto sobre el cual se van a sacar conclusiones, y si requiere del evaluado el desempeño de la habilidad que se está midiendo, en dicha prueba se puede argumentar evidencia de validez relacionada con el contenido (p. 22).

Esto quiere decir que la validez no sólo se remite a los elementos lingüísticos sino también a la tarea evaluativa en cuanto esta posibilita o no el desempeño de la habilidad a evaluar.

La descripción detallada de la tarea evaluativa es una parte indispensable de la rúbrica puesto que asiste al profesor en el momento de comparar el desempeño del estudiante con los criterios de evaluación de los logros permitiendo una interpretación más válida; esta cualidad en la evaluación se denomina transparencia. La transparencia se da a través del grado de detalle en la información dada al evaluado respecto a la prueba (Carroll, 1993, citado por Arias, Maturana & Restrepo, 2012). García-Ros

⁴ Para el ejemplo de rúbrica que se discute en este artículo se definió como constructo teórico la competencia comunicativa (ver Anexo B). Dicho modelo estuvo acorde con la metodología del programa que en general correspondía al enfoque comunicativo en la enseñanza de lenguas —CLT por sus siglas en inglés— (Savignon, 2001). Finalmente, se establecieron los aspectos de la competencia comunicativa a evaluar con base en el programa del nivel correspondiente. Dichos aspectos son los que se encuentran en los descriptores que hacen parte de la escala (ver Anexo A).

⁵ Para un revisión de constructos lingüísticos ver Fulcher & Davidson (2007).

(2011) señala la transparencia como una de las cualidades percibidas por estudiantes y profesores en estudios sobre la utilidad de las rúbricas como instrumentos de evaluación en el contexto universitario (p. 1047). Paralelamente, Arias, Maturana y Restrepo (2012), en el campo de la enseñanza y aprendizaje de lenguas extranjeras, encontraron que la confiabilidad y la viabilidad de una prueba se relacionaban directamente con la transparencia del instrumento y que esta última afectaba positivamente el desempeño.

En su revisión de investigación sobre la validez de la evaluación utilizando rúbricas en educación superior, Reddy y Andrade (2010) reportan la claridad y propiedad del lenguaje utilizado en la rúbrica como una preocupación central para incrementar la validez (p. 445). Por su parte Jonsson y Svingby (2007), previenen sobre el hecho de que la rúbrica *per se* no facilita una interpretación válida del desempeño ya que no provee representatividad del contenido (p. 137). A pesar de las limitaciones y preocupaciones reportadas por estos estudios en términos de validez en experiencias de evaluación utilizando rúbricas, García-Ros (2011) en un estudio de validación de una rúbrica para evaluar habilidades de presentación oral en contextos universitarios reporta:

Los estudiantes ... destacan su validez [de la rúbrica] 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 desde una perspectiva analítica como holística, especialmente entre las valoraciones globales de los proyectos de trabajo —correlación de .89— (p. 1044).

De igual manera, Muñoz, Álvarez, Casas, Gaviria S. y Palacio (2003), en un estudio empírico realizado en Colombia, reportan validez del constructo y confiabilidad entre moderada y alta entre evaluadores en un sistema de evaluación de la producción oral que incluyó el diseño interno de una rúbrica, y de tareas evaluativas auténticas

específicas alineadas con los estándares establecidos en el programa para los diferentes niveles. Con base en los resultados de García-Ros (2011) y Muñoz et al. (2003) podríamos concluir que la validez en un sistema de evaluación en el que se utilicen rúbricas depende de la coherencia entre la escala, las tareas, los estándares y el modelo teórico que enmarca el programa —lo que confirma los resultados de Jonsson y Svingby (2007) si tenemos en cuenta que la representatividad de contenido estaría presente en la tarea evaluativa, la que debe ser consistente con el constructo teórico que sustenta la evaluación.

Puesto que diseñar una tarea evaluativa auténtica (O'Malley & Valdez, 1996) es tan importante como tener una escala de valoración clara y coherente para incrementar la transparencia y validez del proceso evaluativo, concluyo este apartado con algunas consideraciones en lo que se refiere a diseñar dicho tipo de tarea. En la evaluación en leguas extranjeras, diseñar una tarea evaluativa auténtica es tener presente la función lingüística a desarrollar como eje central del constructo alrededor del cual se conectan los otros elementos. Por ejemplo, si se trata de llevar a cabo una descripción, podemos partir de allí para determinar el contexto, quiénes actúan en la situación y cuáles son sus roles; esto determinará otros aspectos de la lengua necesarios en la comunicación para la situación dada.⁶ En general, en la elaboración de *las instrucciones de la tarea* es importante tener en cuenta los siguientes pasos.

- Especificar el género o el formato que enmarca la tarea: una composición corta, un cuento, una presentación, una entrevista, etc.
- Dar detalles sobre el contexto en el que se realiza la tarea: la situación, el lugar, los actores, el rol del estudiante.
- Dar instrucciones precisas en cuanto a los aspectos lingüísticos que espera que el

⁶ Ver en el *Anexo A* los detalles de la tarea evaluativa en la que se pide al estudiante elaborar un perfil escrito a través del cual el profesor podrá evaluar su *competencia comunicativa* a través de la escritura.

estudiante desarrolle: la función lingüística a evaluar: descripción, información, argumentación, etc.; el tema: la familia, la escuela, etc.; los aspectos gramaticales: la sintaxis, el vocabulario, la pronunciación, la ortografía, la puntuación, etc.

- Dar instrucciones precisas respecto a los aspectos mecánicos: títulos, márgenes, etc.
- Especificar el tiempo estimado para que el estudiante complete la tarea y los recursos que puede utilizar.⁷

En resumen, el diseño e implementación de rúbricas, en los términos aquí sugeridos, incrementaría eventualmente la validez en la evaluación en la medida en que: (a) se conecte la tarea evaluativa con el programa del curso: diseñar una rúbrica debería por naturaleza permitirle al profesor evaluar lo que ha enseñado; (b) el constructo a evaluar esté explícito y claro tanto en las instrucciones para el desarrollo de la tarea evaluativa como en los criterios de evaluación descritos en la escala —lo que facilita y valida la valoración del desempeño del estudiante; y (c) el proceso evaluativo se lleve a cabo de forma directa (Lippman, 2003), y a través de una tarea auténtica (O'Malley & Valdez, 1996). Sin embargo, debemos tener presente, los estudios muestran que la validez de contenido y del constructo en experiencias de evaluación con rúbricas no han sido estudiados suficientemente (ver Reddy & Andrade, 2010).

La rúbrica y el impacto.

El impacto de la evaluación debe considerarse en relación con la sociedad, los sistemas educativos y los individuos que componen dichos sistemas. Puesto que las decisiones que influyen en los procesos evaluativos pueden afectar positiva o negativamente a los individuos en sus proyectos de vida, Shohamy (2001) advierte que determinar

el impacto de la evaluación es una responsabilidad de todos los implicados con el fin de prevenir prácticas antidemocráticas. El impacto es entonces una cualidad especialmente significativa en el desarrollo de prácticas evaluativas justas.

El impacto a nivel del aula, denominado *washback*⁸ por algunos autores, se refiere al efecto que la evaluación tiene en el proceso de enseñanza-aprendizaje (Bachman & Palmer, 1996; Brown, 2004). Un impacto positivo está directamente relacionado con las posibilidades de que un evento evaluativo facilite el desarrollo del aprendizaje en los estudiantes, además de medir determinados logros, y con la calidad de la realimentación que el procedimiento proporciona (Brown & Hudson, 1998; Gipps, 1999; Wiliam, 2011). El tipo y continuidad de la realimentación, por ejemplo, permitirá que la evaluación derive o no en el desarrollo de estrategias metacognitivas tales como el auto-monitoreo y la capacidad de autoevaluación (Gipps, 1999; Lamb, 2010; Picón, 2012), e influya positivamente en la auto-regulación y auto-eficacia del estudiante, disminuyendo sentimientos de ansiedad en la evaluación (García-Ros, 2011). Estos aspectos formativos de la evaluación con rúbricas han sido señalados en estudios diversos a nivel de educación básica, secundaria, y superior en general, y en diferentes campos del conocimiento (ver García-Ros 2011; Jonsson & Svingvy, 2007; Muñoz & Álvarez, 2010; Panadero & Jonsson, 2013; Reddy & Andrade, 2010 y Ross, 2006).

En este orden de ideas, Suskie (2002) propone definir y compartir con los estudiantes los resultados esperados de la tarea y la rúbrica con la que su aprendizaje será evaluado como una de las estrategias a seguir en la búsqueda de la justicia (p. 9). Utilizar escalas de valoración consensuadas con los estudiantes, tanto para evaluarlos como para guiar su auto-evaluación, los involucra activamente en el proceso, lo cual arroja resultados positivos:

⁷ Para mayor información acerca de tareas evaluativas auténticas ver O'Malley & Valdez (1996).

⁸ La palabra inglesa *washback* significa efecto de arrastre y se refiere en la evaluación al efecto de esta en el proceso de enseñanza-aprendizaje.

- Los estudiantes internalizan los criterios, lo cual los empodera para argumentar sus juicios en el momento de la evaluación.
- Saben lo que el profesor espera de ellos.
- Encuentran significativas las calificaciones.
- Desarrollan habilidades de argumentación para apoyar sus puntos de vista (Picón, 2007).⁹

En general existen dos tipos de escalas: holísticas y analíticas. Las escalas holísticas presentan una descripción general de los diferentes niveles de logro esperados en el desempeño de los estudiantes respecto a una tarea determinada. Consecuentemente, no se refieren a cada uno de los elementos definidos en el constructo sino que estos se encuentran implícitos en la descripción. Las escalas analíticas en cambio describen cada uno de los principales elementos del marco definido como constructo para cada nivel de logro en la escala.¹⁰

86

Las escalas analíticas son más efectivas en procesos de evaluación en el aula puesto que posibilitan una realimentación más detallada, descriptiva y dialógica entre profesores y alumnos (Gipps, 1999) —lo cual permite llevar a cabo un tipo de evaluación más formativa. El ejemplo de rúbrica que se presenta en este artículo posee una escala analítica que describe cada uno de los elementos del constructo para el nivel excelente y se ha dejado al profesor libertad para decidir los niveles aceptable e inferior a aceptable (ver Anexo A). Aunque este tipo de escala no ofrece detalle para cada uno de los niveles, es de fácil elaboración para un profesor que se inicia en el diseño de estos instrumentos. La experiencia le dará al docente los elementos necesarios para elaborar descripciones más detalladas a medida que se familiariza con los diferentes niveles de desempeño de sus estudiantes.

Resumiendo, la transparencia de la rúbrica unida a la participación de los estudiantes en su diseño e implementación, le permiten tanto a éstos como al docente aprender del evento y desarrollar una evaluación de carácter formativo donde la realimentación se da de forma natural y continua. Así mismo, compartir la rúbrica desde el principio y promover su uso en la autoevaluación y planeación ayudará a que el proceso se lleve de manera sistemática y a que el carácter autorregulador de la evaluación se logre de manera exitosa (Arias, Estrada, Areiza & Restrepo, 2009). Por último, puesto que la participación de los estudiantes en el proceso es primordial para que se dé una evaluación justa, es importante seguir principios democráticos en la evaluación.

La rúbrica y la democracia.

El concepto de justicia que Suskie (2002) promueve está en armonía con el desarrollo de procesos evaluativos democráticos. Además de proponer la *inclusión* de los estudiantes en la definición de los objetivos de la tarea y el diseño de la rúbrica —lo que permite una relación de *poder* más horizontal— sugiere que los profesores evalúen los resultados teniendo en mente las posibilidades de que la enseñanza pudiera no haber sido clara o de que los instrumentos pudieran haber sido inapropiados —lo que eventualmente promovería la *reflexión* y permitiría que los docentes se hicieran *conscientes* de su *responsabilidad* como evaluadores (pp. 9–10).

El principio de democracia implica llevar a cabo prácticas evaluativas dentro de un marco que promueva el empoderamiento de los evaluados, la colaboración y la participación de los involucrados, y el establecimiento de una vigilancia constante para contrarrestar un eventual

⁹ Algunas de estas ideas fueron publicadas en inglés en Picón (2007). Estudios empíricos en el área de las lenguas extranjeras, mencionados en este artículo, quienes también apoyan esta aseveración son Muñoz y Álvarez (2008), Muñoz y Álvarez (2010), Muñoz, Palacio y Escobar (2012) y Picón (2012).

¹⁰ Para mayor información respecto a tipos de rúbricas y su elaboración ver Mertler (2001), Moskal (2000) y Stevens y Levi (2005).

impacto negativo y prevenir usos autocráticos no éticos de las pruebas. Shohamy (2001) asegura que “adoptar principios democráticos ... implica que el acto de evaluar es un esfuerzo mutuo entre evaluadores y evaluados” (p. 379) y propone una serie de estrategias de evaluación basadas en el seguimiento de principios encaminados a que la sociedad pueda protegerse de prácticas evaluativas antidemocráticas. Entre ellas menciona prácticas que promuevan la colaboración y participación de quienes van a ser evaluados en el proceso de evaluación y formas de promover el compromiso y la responsabilidad por las pruebas y sus usos en todos aquellos involucrados en el acto evaluativo.

En este orden de ideas, Hewitt (1995) defiende lo que llama “el diseño local de estándares” por considerar que son más motivadores para los estudiantes y pueden ser desarrollados en el aula de clase o por entes administrativos de la escuela. El autor enfatiza que cuando los estudiantes participan en el establecimiento de criterios de evaluación, estos se constituyen en un set de objetivos a ser alcanzados por ellos mismos que además pueden ser reevaluados a lo largo del año escolar. Hewitt asegura además que para los estudiantes que aprenden a evaluar su trabajo bajo criterios que ellos mismos han desarrollado, es más fácil evaluarlo con base en estándares externos. El autor hace énfasis en la importancia de definir muy bien los criterios de evaluación de manera que todos los implicados en el proceso los entiendan y sugiere los siguientes pasos para lograrlo:

- Hacer una lluvia de ideas con los estudiantes acerca de los criterios necesarios para evaluar un aspecto dado.
- Discutir cuáles de ellos serían esenciales en la evaluación.

- Mejorar la redacción, reducir el número de criterios de forma que sean manejables, y decidir cómo los estándares serán aplicados y cómo describir el nivel de logro de los estudiantes.
- Desarrollar escalas analíticas para cada estándar y/o una escala de valoración holística para combinarla con las analíticas.

En la evaluación de un constructo complejo como la *competencia comunicativa* (Anexo B), es necesario guiar la discusión con los estudiantes con base en un formato tentativo, o un grupo de criterios, que el docente haya diseñado previamente. Puesto que este es el caso que se presenta como ejemplo central en esta reflexión, sugiero seguir los siguientes procedimientos:

- Pensar en los criterios de evaluación con base en los contenidos que se van a evaluar y los objetivos que se espera que los estudiantes logren.
- Recordar que es necesario tener claro el constructo. Para el caso de la *competencia comunicativa*, tener en mente los diferentes componentes de dicho modelo en relación con los contenidos a evaluar en el momento de definir el marco (ver Anexo B).
- Si el programa de curso explicita estándares de desempeño, estos pueden servir como descriptores en la rúbrica.
- Se deben expresar en forma sencilla los objetivos que se espera que los estudiantes alcancen, al igual que el nivel de desempeño esperado para que un estudiante logre cada uno de ellos.¹¹

¹¹ La descripción del desempeño esperado en un aspecto determinado será un criterio para evaluarlo. Por ejemplo, si se quiere evaluar el desempeño de un estudiante en su escritura en lengua extranjera y los contenidos del programa incluyen la descripción de la familia, uno de los criterios de evaluación sería que *el estudiante describe un determinado número de miembros*, otro que *utiliza adjetivos relacionados con la apariencia física y la personalidad*, y otro más sería que *forma oraciones sencillas en presente simple*.

Una vez elaborada la rúbrica tentativa sugiero seguir los siguientes pasos en la clase:

- Presentar los criterios en un formato que sea claro para los estudiantes y darles tiempo para discutirlos en grupos pequeños.
- Utilizar la lengua materna y un lenguaje simple.
- Animarlos a expresar cualquier desacuerdo o ambigüedad que pudieran encontrar.
- Pedirles que escriban, en el mismo formato, sus ideas y opiniones respecto a estos criterios de manera que el profesor pueda organizar la rúbrica para la siguiente clase.
- Llevarse la realimentación y organizar la rúbrica definitiva incluyendo las ideas de los estudiantes.
- Presentar el formato definitivo en la clase siguiente y discutir la escala de valoración.

88

Si bien a lo largo del artículo se hace referencia a estudios empíricos y revisiones de literatura sobre la validez del uso de las rúbricas, quisiera enfatizar en el aspecto formativo que este instrumento posibilita. Aún más, la oportunidad que presenta el diseño de rúbricas para promover procesos democráticos en el aula y seguir principios de justicia en la evaluación, y que de manera consciente subrayo a lo largo de esta discusión, hace de esta reflexión un esfuerzo por promover prácticas evaluativas insertadas en enfoques críticos en la enseñanza y aprendizaje de lenguas (Gipps, 1999).

Para concluir quisiera dejar mi testimonio de que, a pesar del esfuerzo que requiere la elaboración de rúbricas, en realidad vale la pena diseñarlas y hacerlo de manera consensuada. Además de las cualidades ya discutidas ciertamente le ahorran a los profesores, no solo tiempo, sino también sentimientos de inseguridad y ansiedad en el momento de la

calificación. Más importante aún, el trabajo paga en términos de aprendizaje y justicia tanto para docentes como alumnos.

Conclusiones

En esta reflexión se ha discutido la utilización de rúbricas como instrumento de evaluación alternativa que posibilita procedimientos justos —es decir equitativos, válidos, transparentes, formativos y enmarcados en procesos democráticos. Consecuentemente se ha argumentado que el diseño consensuado y aplicación de rúbricas en la evaluación, en las condiciones aquí definidas, posibilita el desarrollo de la justicia ya que: (a) se ajusta al concepto de *perspectiva múltiple* facilitando procedimientos equitativos; (b) permite aumentar la validez de las pruebas en cuanto exige una definición clara del constructo, se ajusta al diseño de tareas evaluativas auténticas variadas y facilita una interpretación pertinente de los resultados; (c) incrementa la transparencia en la evaluación al ofrecer instrucciones detalladas de la tarea evaluativa y una escala de valoración analítica; (d) permite llevar a cabo evaluación formativa debido a la calidad de realimentación que facilita; y (e) proporciona espacios para la participación de los estudiantes en el proceso.

Las ideas aquí presentadas aplican para el desarrollo e implementación de rúbricas por profesores en el aula de clase y nacen de experiencias analizadas en tal contexto. Sin embargo se deja claro que entender y definir un constructo lingüístico desde su dimensión teórica, así como diseñar un sistema de calificación e interpretación coherente con dicho constructo, requiere un estudio juicioso y/o el apoyo de expertos. En ese sentido, algunos estudios mencionados han mostrado que los grupos de estudio de profesores y la investigación acción participativa en el aula son estrategias pertinentes y efectivas en el desarrollo profesional de los docentes que les permiten ganar experticia y autonomía técnica (ver Arias, Maturana & Restrepo, 2012; Birchak et al., 1998; Frodden & Picón, 2005; Picón, 2012).

Implicaciones y sugerencias para investigación

Es importante llegar a consenso entre coordinadores y profesores respecto a los estándares de contenido y desempeño esperados para unificar criterios de evaluación en programas de lenguas extranjeras. Es igualmente significativo posibilitar espacios de estudio —tanto para los primeros como para los últimos— en torno a los temas que se tratan en este artículo y a los modelos de lengua que definen enfoques de enseñanza y aprendizaje en el campo de las lenguas extranjeras. La experiencia que enmarca esta reflexión se inició en un proyecto de investigación fruto de un grupo de estudio que reúne coordinadores y profesores de programas de inglés como lengua extranjera de la Escuela de Idiomas de la Universidad de Antioquia y es por ello que recomiendo esta forma de desarrollo profesional.

Es indispensable que los profesores tengan apoyo en el diseño e implementación de las tareas evaluativas y las escalas de valoración que van a utilizarse para calificar el desempeño de los estudiantes, y que tengan autonomía para adaptarlas. De esta manera los profesores se sienten seguros y comprometidos a enseñar lo que institucionalmente se ha definido en el programa al tiempo que la validez se incrementa.¹² A este respecto, Muñoz y Álvarez (2010) y Muñoz, Palacio y Escobar (2012) enfatizan en la necesidad

de (a) comprender las creencias de los profesores hacia la evaluación y (b) brindarles apoyo constante en sus prácticas evaluativas para lograr un impacto positivo en la enseñanza.

Estudios investigativos sobre la confiabilidad y validez de procedimientos de evaluación alternativos utilizando rúbricas elaboradas con la participación de alumnos, docentes, administradores y expertos, en contextos de enseñanza de las lenguas extranjeras serían un valioso aporte a la discusión en este campo. Es igualmente significativo estudiar constructos de evaluación en enfoques contemporáneos de enseñanza y aprendizaje de lenguas con el fin de posibilitar el diseño de rúbricas coherentes y válidas para dichos enfoques.

Finalmente, se sugiere tomar en cuenta las políticas lingüísticas establecidas en el contexto nacional e internacional en términos de estándares de evaluación de la suficiencia en lengua extranjera, con el fin de conectar escalas de evaluación del desempeño de los alumnos en el aula y escalas de suficiencia utilizadas en el ámbito internacional tales como las propuestas por el Marco Común Europeo de Referencia (ver Instituto Cervantes 2002).¹³

¹² El Sistema de Evaluación en Lenguas Extranjeras elaborado por Arias et al. (2009) es un texto pertinente para un programa de desarrollo profesional en el tema de la evaluación de lenguas extranjeras.

¹³ Como referencia en la creación participativa de escalas contextualizadas para la evaluación de la suficiencia ver Halonnen et al. 2009)

REFERENCIAS

- Arias, C., Estrada, L., Areiza, H., & Restrepo, E. A. (2009). *Sistema de evaluación en Lenguas Extranjeras*. Medellín, Colombia: Reimpresos, Universidad de Antioquia.
- Arias, C., & Maturana, L. (2005). Evaluación en lenguas extranjeras: Discursos y prácticas. *Íkala, revista de lenguaje y cultura*, 10, 63–91.
- Arias, C., Maturana, L., & Restrepo, M. I. (2012). Evaluación de los aprendizajes en lenguas extranjeras: Hacia prácticas justas y democráticas. *Revista Lenguaje*, 40(1), 99–126.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Nueva York, NY: Oxford University Press.
- Birchak, B., Connor, C., Crawford, K. M., Kahn, L.H., Kaser, S., Turner, S., & Short, K. (1998). *Teacher study groups: Building community through dialogue and reflection*. Urbana, Ill: National Council of Teachers of English.
- Brown, D. H. (2004). *Language assessment: Principles and classroom practices*. New York, NY: Longman.
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653–675.
- Frodden, C., & Picón, E. (2005). El desarrollo de la autonomía del profesor: Un estudio de caso. *Boletín de Investigación Educativa*, 20, 285–300.
- Frodden, C., Restrepo, M. I., & Maturana, L. M. (2004). Analysis of assessment instruments used in foreign language teaching. *Íkala, revista de lenguaje y cultura*, 9(15), 171–201.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. New York, NY: Routledge Applied Linguistics.
- García-Ros, R. (2011). Análisis y validación de una rúbrica para evaluar habilidades de presentación oral en contextos universitarios. *Electronic Journal of Research in Educational Psychology*, 9, 1043–1062.
- Genesee, F., & Upshur, J. (1996). *Classroom-based evaluation in second language education*. New York, NY: Cambridge University Press.
- Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education*, 24, 355–392.
- Goodrich-Andrade, H. (n.d.). *Understanding rubrics*. Recuperado de <http://learnweb.harvard.edu/alps/thinking/docs/rubricar.htm>
- Halonnen, J. S., Bosack, T., Clay, Sh., McCarthy, M., Dunn, D. S., Hill, G. W., McEntarffer, R., Mehrotra, Ch., Nesmith, R., Weaber, K. A., & Whitlock, K. (2009). A rubric for learning, teaching, and assessing scientific inquiry in Psychology. *Teaching of Psychology*, 30, 196–208. doi:10.1207/S15328023TOP3003_01
- Hewitt, G. (1995). *A portfolio primer: Teaching, collecting, and assessing student writing*. Portsmouth, NH: Heinemann.
- Instituto Cervantes (2002). *Marco de referencia europeo para el aprendizaje, la enseñanza y la evaluación de lenguas* [documento electrónico], Estrasburgo, (Traducción al español de Alejandro Valero Fernández). Recuperado de <http://cvc.cervantes.es/obref/marco/>
- Instituto de Estudios sobre Desarrollo y Cooperación Internacional, HEGOIA (n.d.). *Equidad*. Recuperado de <http://www.bantaba.ehu.es/obs/ocont/eq/>
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130–144. Retrieved from <http://dx.doi.org/10.1016%2Fj.edurev.2007.05.002>
- Lamb, T. (2010). Assessment of autonomy or assessment for autonomy? Evaluating learner autonomy for formative purposes. En A. Paran & L. Sercu (Eds.). *Testing the untestable in language education* (pp. 98–119). Bristol, UK: Multilingual Matters.
- Lippman, J. (2003). Assessing writing. En I. Clark & B. Bamberg (Eds.), *Concepts in composition: Theory and practice in the teaching of writing* (pp. 199–220). Mahwah, NJ: Erlbaum. Tomado de University of Pittsburg NETLIBRARY database (79439) http://www.library.pitt.edu/articles/database_info/netlibrary.html
- Lopez, A., & Bernal, R. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *PROFILE, Issues in Teachers' Professional Development*, 11(2), 55–70.
- Mansoor, I., & Grant, S. (2002). A writing rubric to assess ESL student performance [Electronic version]. *Adventures in Assessment*, 14, 33–38. (ERIC Document Reproduction Service No. ED 482885).

- Mertler, Craig A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7(25). Tomado de <http://PAREonline.net/getvn.asp?v=7&n=25>
- Moskal, Barbara M. (2000). Scoring rubrics: what, when and how. *Practical Assessment, Research & Evaluation*, 7(3). Tomado de <http://PAREonline.net/getvn.asp?v=7&n=3>
- Muñoz, A., & Álvarez, M. (2008). A study of the effect of a writing assessment system on instruction. *MEXTESOL Journal*, 32(2), 55–70.
- Muñoz, A., & Álvarez, M. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27(1), 1–17.
- Muñoz, A., Álvarez, M., Casas, S., Gaviria S., & Palacio, M. (2003). Validation of an oral assessment tool for classroom use. *Colombian Applied Linguistic Journal*, 137–157
- Muñoz, A., Palacio, M., & Escobar, L. (2012). Teachers' beliefs about assessment in an EFL context in Colombia. *Profile*, 14(1), 143–158.
- O'Malley, J. M., & Valdez, P. L. (1996). *Authentic assessment for English language learners: Practical approaches for teachers*. New York, NY: Longman.
- Oxford, R. (2001). Language learning styles and strategies. En M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (3rd ed.), (pp. 359–366). Boston, MA: Heinle & Heinle.
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144. <http://dx.doi.org/10.1016/j.edurev.2013.01.002>
- Picón, E. (2007). Rubric: A scoring device to enhance fair assessment. *Cartilla pedagógica para la formación de maestros en ejercicio*. Medellín, Colombia: Escuela de Idiomas, Universidad de Antioquia.
- Picón, E. (2012). Promoting learner autonomy through teacher-student partnership assessment in an American high school: A cycle of action research. *PROFILE, Issues in Teachers' Professional Development*, 14(3), 145–162.
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment and Evaluation in Higher Education*, 35, 435–448. doi:10.1080/02602930902862859
- Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research & Evaluation*, 11(10), 1–13.
- Savignon, S. J. (2001). Communicative language teaching for the twenty-first century. M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (3rd ed.), (pp. 13–28). Boston, MA: Heinle & Heinle.
- Shohamy, E. (1998). Evaluation of learning outcomes in second language acquisition: A multiplism perspective. En Byrnes H. (Ed.), *Learning foreign and second languages: Perspectives in research and scholarship* (pp. 238–261).
- Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing* 18(4), 373–391.
- Shohamy, E., & Inbar, O. (2006). *The language assessment process: A multiplism perspective*, (Professional Development Document 0603). University Park, PA: The Pennsylvania State University, Center for Advanced Language Proficiency Education and Research.
- Stevens, D., & Levi, A. (2005). *Introduction to rubrics: An assessment tool to save grading time, convey effective feedback and promote student learning*. Sterling, VA: Stylus.
- Suskie, L. (2002). Fair assessment practices: Giving students equitable opportunities to demonstrate learning [Electronic version]. *Adventures in Assessment*, 14, 5–10. (ERIC Document Reproduction Service No. ED 482885).
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37, 3–14.
- Williams, J. (2003). *Preparing to teach writing: Research, theory and practice* (3rd ed.). Mahwah, NJ: Erlbaum. Tomado de University of Pittsburg NETLIBRARY database (83850), http://www.library.pitt.edu/articles/database_info/netlibrary.html

How to reference this article: Picón, E. (2013). La rúbrica y la justicia en la evaluación. *Íkala, revista de lenguaje y cultura*, 18(3), 79–94.

ANEXO A

Rúbrica

UNIVERSIDAD DE ANTIOQUIA
ESCUELA DE IDIOMAS
CENTRO DE EXTENSIÓN
PROGRAMA DE INGLÉS PARA JOVENES

LEVEL 1

Written task: Profile
Evaluación sumativa (____ %)

92

SITUATION: You are registered in a level one of a teenagers' English program and your teacher wants you to exchange profiles with a student from another level-one course. The activity is intended for you to get to know each other and become pen pals for the rest of the course.

TASK: A profile is a short description of a person. It may include personal information such as his or her origin, age, personality, hobbies, etc. Write your profile following the instructions below. You can decide if you will send it in a letter or by email. Do your best to cause a good impression to your new friend. You have 30 minutes to do this activity.

To complete this task successfully you need to:

- Write three five-line paragraphs introducing yourself to your classmate. Include some personal information, your daily routine and some likes and dislikes.
- Use the simple present tense in affirmative and negative forms using the verb to be and others such as live, study, like, dislike, love, hate, enjoy, and want.
- Tell your new pen pal how often you do each activity when describing your daily routine. Use time and frequency words.
- Take into account the appropriate use of commas, periods, capital letters and spelling.
- Use the checklist given at the end to revise your work.

Tabla A1

PASSING SCORE: 3.0

Student's Name:		Date:					
%	Criteria / Score (number 1 is the lowest scale; 5, 15, 20 and 25 the highest)	Poor	Needs Improvement	Fair	Good	Outstanding	Total
	The student is able to:						
25	Provide personal information, describe daily routine and mention likes and dislikes.	1–5	6–10	11–15	16–20	21–25	
25	Express ideas in a clear, connected and logical form.	1–5	6–10	11–15	16–20	21–25	
20	Use some structures worked in the course such as: adjective-noun word order, to be, simple present tense in affirmative and negative form.	1–4	5–8	9–12	13–16	17–20	
15	Use appropriate vocabulary to express ideas about personal information, daily routine and likes and dislikes.	1–3	4–6	7–9	10–12	13–15	
5	Use of punctuation in an appropriate way (commas and periods).	1	2	3	4	5	
5	Use capital letters.	1	2	3	4	5	
5	Have a spelling of words without interfering with meaning.	1	2	3	4	5	
Comments:						Final Score	
						_____ /2 =	

93

Nota: Preparado por miembros del Grupo de Estudio de Rúbricas, Escuela de Idiomas, Universidad de Antioquia, Septiembre 2009.

Anexo B

Constructo

Tabla B1

<p>Competencia Sociocultural</p> <p>Reglas sociales de uso de la lengua, roles de los participantes, información compartida, funciones de la interacción, convenciones, pertinencia, conciencia cultural</p>		
<p>Competencia Gramatical</p> <p>Lexico, Morfología, Sintáxis, Fonología</p>	<p>Competencia Discursiva</p> <p>Coherencia y Cohesión</p>	<p>Competencia Estratégica</p> <p>Pedir información, buscar aclaración, negociar significado, parafrasear, etc.</p>

Nota: Competencia comunicativa, tomado de Savignon (2011).