

Experiencia de Diseño y Uso de una Rúbrica para Evaluar Informes de Laboratorio en Formato Publicación Científica

Carola E. Bruna⁽¹⁾, Verónica A. Villarroel⁽²⁾, Daniela V. Bruna⁽²⁾ y José A. Martínez⁽¹⁾

(1) Universidad de Concepción, Facultad de Ciencias Biológicas, Departamento de Bioquímica y Biología Molecular. Edmundo Larenas 1290, Concepción, Chile. (e-mail: carolabruna@udec.cl; jmartine@udec.cl)

(2) Universidad del Desarrollo. Facultad de Psicología. Centro de Investigación y Mejoramiento de la Educación (CIME). Ainavillo 456, Concepción, Chile. (e-mail: vvillarroel@udd.cl; dbrunaj@udd.cl)

Recibido Jun. 11, 2018; Aceptado Ago. 7, 2018; Versión final Oct. 5, 2018, Publicado Abr. 2019

Resumen

En este trabajo se diseñó una intervención para contribuir a la formación de científicos, a través de rediseñar la evaluación de una tarea de desempeño auténtica. Para ello, se creó y validó una rúbrica analítica para evaluar un informe de laboratorio en formato publicación científica en la asignatura bioquímica. Se obtuvo un consenso entre evaluadores de 70%. El análisis permitió mejorar el instrumento e identificar debilidades en el desempeño de los estudiantes. Se obtuvo la percepción de los estudiantes de la experiencia, quienes declararon utilizar la rúbrica, apreciándose acuerdo respecto a su calidad, objetividad y pertinencia. Sin embargo, no hubo acuerdo respecto de la utilidad para guiar y retroalimentar, ni hubo efecto en el rendimiento, evidenciando que se requiere robustecer el diseño. En consecuencia, se generó una propuesta que incorpora: involucrar a los estudiantes en la definición de estándares de calidad, guiar el análisis de datos, incluir coevaluación y retroalimentación.

Palabras clave: evaluación; rúbrica; retroalimentación; informe de laboratorios; publicación científica

Experience of Design and Use of a Rubric to Assess Laboratory Reports Formatted as Scientific Publications

Abstract

In this work, an intervention was designed to contribute to the training of scientists, through redesigning the assessment of an authentic performance task. An analytic rubric to assess a laboratory report written as a scientific paper for a biochemistry course was created and validated. The agreement between referees was 70%. The analysis allowed improving the instrument and identifying weaknesses in the students' performance. Students' perception of this experience was obtained, who declared that they used the rubric, agreeing in regard of its quality, objectivity and pertinence. However, there was no agreement with respect of its usefulness for guiding and providing feedback. Also, no effect was found on the course grades, showing that it is necessary to strengthen the design. As a result, a proposal was generated including: involving students in the definition of the quality standards, guiding the data analysis, including peer assessment and feedback. This experience illustrates the need of continuously updating and improving the teaching practices in order to achieve the learning outcomes.

Keywords: assessment; rubric; feedback; laboratory report; scientific paper

INTRODUCCIÓN

Actualmente es ampliamente aceptado que la enseñanza en educación superior debe centrarse en el estudiante y enfocarse al logro de competencias profesionales específicas y genéricas en lugar de orientarse a la transmisión de información (Barr y Tagg, 1995; Biggs 1999; Figueroa, Gilio y Gutiérrez, 2008; Villarroel y Bruna, 2014). Para ello, el docente debe considerar cómo aprende particularmente el grupo de estudiantes a los que enseña, reflexionar sobre cómo enseñar, identificar qué contenidos y competencias intencionar, basándose en el currículum y perfil de egreso. Asimismo, debe diseñar las actividades e instrumentos para evaluar los aprendizajes en concordancia con las estrategias de enseñanza utilizadas (Harden y Crosby, 2000; Pales, 2006). En algunos países, las instituciones de educación superior han generado instancias para que los docentes actualicen sus metodologías de enseñanza, para conocer el perfil de los estudiantes que ingresan y actualizar los planes curriculares, por ejemplo a través de la formación de centros de apoyo a la docencia, programas estatales orientados a mejorar la calidad de la docencia, fondos internos concursables para el mejoramiento docente, evaluación y premios a la excelencia docente (González, 2015). Sin embargo, la evaluación es el ámbito en que menos se ha progresado (Boud y Falchikov, 2007).

Se ha reportado que la forma en cómo se lleva a cabo la evaluación tiene implicancias en el aprendizaje de los estudiantes. Se ha reportado que aquellos profesores que perciben la enseñanza como transmisión de conocimiento, perciben la evaluación como un elemento separado de la enseñanza, en lugar de una estrategia integrada que promueve el aprendizaje profundo. Por otra parte, aquellos que perciben la enseñanza como la construcción de conocimiento, adoptan diferentes prácticas de evaluación, además de las tradicionales. Esto es relevante, ya que los métodos de evaluación que utilizan los profesores universitarios tienen un rol importante en la calidad del aprendizaje (Ribeiro y Assunção, 2016). En este contexto, en las últimas dos décadas, los investigadores y administradores han propiciado un cambio en la cultura de la evaluación que ha sido reconocido en educación superior (Sambell, 2016). No obstante, a pesar de haber logrado progresos, en la práctica, la evaluación ha sido particularmente resistente al cambio, permaneciendo como el ámbito más conservador (Bloxham, 2016).

En muchos países, las pruebas de lápiz y papel son las más comúnmente utilizadas, las cuales no son indicadores de habilidades como resolución de problemas, razonamiento o pensamiento crítico. El problema de este tipo de evaluaciones es que no miden competencias ni estimulan el desarrollo de éstas (Villarroel et al., 2015). Tampoco proveen información respecto de los hábitos de trabajo o habilidad para cooperar, siendo más relevantes para el profesor, que para los estudiantes (Vanaki y Memarian, 2009). Si bien pueden ser efectivas en algunos contextos y para algunos propósitos, hay evidencias que indican que promueven bajo nivel de comprensión, estimulando la reproducción y memorización (Ribeiro y Assunção, 2016). Esto expone la necesidad de evaluar el desarrollo de niveles cognitivos superiores y logro de competencias mediante otros instrumentos. Sin embargo, se ha reportado que al no centrarse en lo memorístico y disciplinar, las estrategias e instrumentos parecen subjetivos, siendo la evaluación percibida como injusta por los estudiantes (Allen y Tanner, 2006). Por otra parte, el aprendizaje se ha sacrificado para cubrir contenidos. El número de estudiantes, la extensión del currículum y la sobrecarga de los profesores son problemas que han contribuido a que grandes cambios en la educación superior no se generen, especialmente en el ámbito de la evaluación. (Boud y Molloy, 2013). Es necesario repensar y cuestionarse si la forma en que se evalúa el conocimiento está logrando que los estudiantes aprendan de manera significativa y profunda (Ohaja et al., 2013).

Actualmente, se concibe a la evaluación como un proceso dinámico y continuo en la cual se obtiene información para emitir juicios de valor y tomar decisiones. Se espera que no se diseñe solamente como una instancia final para calificar (Reddy y Andrade, 2010; Villarroel et al., 2017a). La evaluación representa un desafío tanto para los estudiantes, como para el cuerpo de profesores. Se requiere adoptar estrategias que hayan sido investigadas, ya que actualmente hay una discordancia entre el aprendizaje y la evaluación, que tiene implicancias en la integridad del proceso educativo (Schell y Porter, 2018). Un enfoque para avanzar, es el paradigma “evaluación como aprendizaje”, en el cual la evaluación se concibe como una instancia cuya resolución es aprender (Earl, 2013). Otra propuesta en este ámbito, es acercar el mundo del trabajo al aula, a través contextualizar las evaluaciones e incluir tareas realistas en términos de proceso y productos, lo cual se plantea en el modelo de Evaluación Auténtica (Palm, 2008; Villarroel et al., 2018).

En la evaluación auténtica las condiciones de la evaluación son presentadas en contexto, se mantienen reales. Al replicar los desafíos y estándares a los que los estudiantes se enfrentarán en el mundo del trabajo, tiene un valor más allá de la universidad, el currículum y las prácticas de aula de aprendizaje y enseñanza. (Palm, 2008). En el ámbito de la enseñanza de las ciencias, le provee al profesor la oportunidad de introducir nuevas metodologías instruccionales que favorecen la investigación y la creación de conocimiento (Vanaki y Memarian, 2009). Experiencias auténticas de investigación en el aula son consideradas elementos valiosos para promover la ciencia en el pregrado, motivando al estudiante y vinculando la investigación con la enseñanza (Wilson et al., 2016).

En este modelo, es un desafío incluir tareas basadas en desempeño, refiriéndose a aquellos instrumentos en los que se simulan los procesos y productos profesionales, en los cuales los estudiantes deben mostrar un desempeño que involucre construcción de conocimiento y desarrollo de habilidades cognitivas superiores (Rudner y Boston, 1994), permitiendo evaluar competencias que correspondan con las actividades y problemas disciplinares en el campo laboral, avanzando en contribuir al alineamiento entre la enseñanza y la evaluación y otorgando autenticidad (Martínez-Rizo y Mercado, 2015).

El desarrollo de competencias científicas constituye un eje estratégico del desarrollo humano, ya que implica la promoción de importantes habilidades, como lo son la capacidad crítica de la sociedad (Alberts, 2008) y la indagación científica, entendida como la habilidad de plantear preguntas acerca del mundo natural, generar hipótesis, diseñar y ejecutar un proceso de investigación para poder responderlas (Winschitl, 2003). Las disciplinas científicas deben ser enseñadas con métodos activos, enfocados en la aplicación (Waldrop, 2015), en parte a través de métodos experimentales para contribuir a la adquisición de competencias procedimentales, resolución de problemas, desempeño académico y a la interacción social (Ifeanyichukwu, 2016). Por lo tanto, y en consecuencia con el modelo de evaluación auténtica, consideramos que en la labor de contribuir a la formación de científicos, se debe incorporar en el aula algunas de las tareas que enfrentará el futuro profesional, tales como: diseño de experimentos, análisis de datos, ejecución de experimentos, comunicación de resultados en congresos científicos, elaboración de informes de laboratorio, redacción de publicaciones científicas y proyectos de investigación. Algunas de estas actividades son en general incluidas en los programas de asignatura en el pregrado, sin embargo suelen poseer una baja ponderación en la calificación final. Son evaluadas aplicando criterios personales no explicitados ni sistematizados y/o no se provee de instancias para retroalimentar y promover el logro del aprendizaje, usualmente desarrollándose sin involucrar al estudiante en el proceso de evaluación (Stiggins, 2002). Es decir, se requieren más estudios para determinar cómo evaluar el desempeño en el laboratorio (Ifeanyichukwu, 2016).

Si bien existe una alta valoración de los trabajos prácticos, asociándose a un incremento de interés y emoción de los alumnos, no se aplican estrategias para ayudar al alumnado a consolidar las habilidades de razonamiento (Domenech, 2013). También se ha planteado una inconsistencia entre el aprendizaje intencionado y las prácticas evaluativas en proyectos de investigación científica (Wilson et al., 2016), evidenciando falta de claridad en el diseño y la dificultad de diseñar instrumentos para evaluar objetivamente la tarea que a la vez guíen al aprendiz en el proceso. Un instrumento que permite explicitar estándares y recolectar evidencia de los logros de los estudiantes respecto de competencias como comunicación, argumentación, resolución de problemas al desarrollar tareas complejas, es la rúbrica analítica de desempeño (Allen y Tanner, 2006).

La rúbrica como instrumento de evaluación

La rúbrica analítica es un instrumento de evaluación educativa que contempla una escala progresiva de niveles de desempeño que corresponden a rangos de ejecución de una tarea, permitiendo emitir juicios de calidad respecto del trabajo de los estudiantes. Incluye aspectos a evaluar globalizados en dimensiones y criterios que describen clara y explícitamente cada uno de los niveles de desempeño para cada dimensión (Allen y Tanner, 2006). Ha sido utilizada en una amplia variedad de disciplinas. Allen y Tanner (2006) plantean que promueve procesos reflexivos en ambos, profesores y estudiantes. Se plantea útil no sólo para fines evaluativos, sino también como un elemento instruccional, ya que al ser utilizada por los estudiantes para evaluación formativa de su trabajo en progreso, la rúbrica puede enseñar y no sólo evaluar.

El instrumento tiene el potencial de ayudar a los estudiantes a entender qué deben aprender y los estándares de calidad de la actividad asignada (Nicol y Macfallane, 2006; Raposo y Martínez, 2011; Reddy y Andrade, 2010). La percepción del estudiante es que la evaluación es más transparente y justa, que clarifica los objetivos reduciendo la incerteza al aclararles aspectos críticos de la tarea, permitiéndoles regular sus procesos y obtener retroalimentación inmediata, asociándose con una disminución de la ansiedad y con satisfacción con la evaluación (Reddy y Andrade, 2010). Se ha planteado que incluir rúbricas en la evaluación aumenta la autenticidad y motivación. La perspectiva del profesor es que hay mayor transparencia, que promueve la práctica reflexiva y ayuda a retroalimentar (Jonsson y Svingby, 2007). Por otra parte, también se ha planteado útil para explicitar los criterios y guiar la calificación en docentes novatos (Allen y Tanner, 2006).

Prins et al. (2015) plantean los siguientes usos para la rúbrica: 1) Herramienta de evaluación para los profesores; 2) Explicitar expectativas de desempeño a los estudiantes; 3) Familiarizar a los estudiantes con criterios de un buen desempeño en la tarea; 4) Retroalimentación; 5) Auto y co-evaluación. Plantean que la rúbrica puede aumentar la transparencia, reducir la ansiedad, ayudar a retroalimentar, mejorar la autoeficacia y soportar la autorregulación. También discuten que se puede utilizar para seguir metas curriculares.

Considerando lo anterior, en este trabajo se describe una experiencia para mejorar la docencia en el ámbito de la enseñanza de las ciencias. Se la creó y validó una rúbrica analítica de desempeño para evaluar informes de laboratorio en formato de una publicación científica, la cual fue utilizada en la asignatura bioquímica para la carrera de pregrado Bioingeniería. Para el estudiante, esta actividad de evaluación lo acercaría a campo laboral al involucrar dos tareas que son actividades habituales del científico, por otra parte, el instrumento podría guiarlos y retroalimentarlos al disponer desde un inicio de descripciones de lo que se considera un buen desempeño. Para el profesor, la rúbrica podría contribuir a la evaluación objetiva y a la toma de decisiones al identificar fortalezas y debilidades en los logros de los estudiantes.

METODOLOGÍA

Se diseñó y validó un instrumento para evaluar informes de laboratorio en formato publicación científica. Utilizando este instrumento, se llevó a cabo una intervención didáctica para promover competencias científicas asociadas al trabajo de laboratorio y su comunicación en la asignatura bioquímica para la carrera Bioingeniería de la Universidad de Concepción, la cual se dicta el quinto semestre del plan curricular. La asignatura incluye seis sesiones de tres horas de laboratorio, en el cual los estudiantes trabajando en parejas con la guía de un manual de laboratorio, analizan una enzima realizando las etapas en el orden en que esto se llevaría a cabo en una investigación científica. Se realiza la cuantificación, identificación, análisis de pureza, caracterización cinética y el estudio de los factores que afectan la actividad enzimática de la enzima ferredoxina NADP⁺ reductasa basado en Bess et al. (2003). El laboratorio completo se evalúa con un solo informe en formato de publicación científica. Hasta esta experiencia, la calificación, que ponderaba un 12% de la nota final, se obtenía a través de la evaluación de un profesor con experiencia en investigación científica, quien aplicaba criterios personales basados en su práctica como investigador, los cuales no estaban sistematizados ni eran informados detalladamente con anterioridad a los estudiantes.

Elaboración y validación de la rúbrica

Atendiendo a la necesidad de proveer un instrumento para evaluar el informe en formato publicación científica que contribuyera a la objetividad y a guiar a los estudiantes, se elaboró una guía sucinta que describe la estructura del informe y se diseñó una rúbrica analítica de desempeño basándose en la estructura del laboratorio y en las características y requisitos de una publicación científica. Dicha rúbrica contó con 9 dimensiones y 4 niveles de desempeño (versión 1). La rúbrica se diseñó orientada al usuario, de modo que tiene por foco entregar información útil para ayudar al estudiante a entender la retroalimentación y la evaluación (Prins et al., 2015).

Cinco de las nueve dimensiones correspondían a las secciones de una publicación científica (resumen, introducción, métodos, resultados y discusión), dos estaban relacionadas con la estructura (formato, y ortografía y gramática), una con la responsabilidad y finalmente, se incorporó una dimensión titulada apreciación global. Para redactar los criterios se basó en la experiencia de los autores que realizan investigación en ciencias biológicas y que habitualmente realizan la tarea de redactar informes y publicaciones científicas, apoyándose en reportado por Cargill y O'Connor (2009). Se propuso una valoración diferente a algunas las dimensiones otorgando mayor ponderación a los resultados y discusión, basándose en su importancia (Şanlı et al., 2013). El presentar manuscritos para publicación requiere seguir un formato y apropiada redacción, así también, respetar plazos y dedicación para cumplir con la tarea. Estos aspectos fueron incluidos en las dimensiones que no correspondían a secciones de una publicación científica. La dimensión apreciación global estaba relacionada con la valoración del trabajo completo, simulando la evaluación final de un revisor.

El instrumento se validó por juicio de expertos. Para ello se utilizó una lista de cotejo con 10 aspectos a evaluar y espacio para comentarios y sugerencias. 7 docentes expertos en la elaboración de rúbricas y/o en el trabajo científico evaluaron el instrumento. Para estimar la concordancia entre los jueces que evaluaron la Rúbrica, se utilizó Kappa de Fleiss, que permite trabajar con más de dos jueces. Posteriormente el instrumento fue modificado considerando el grado de acuerdo entre jueces y las sugerencias recibidas (versión 2).

Diseño de la intervención

Este estudio se realizó mediante un enfoque cuantitativo, con un diseño de tipo pre-experimental y alcance explicativo. Participaron 15 estudiantes universitarios, 7 (47%) hombres y 8 (53%) correspondientes a la totalidad de los estudiantes cursando la asignatura bioquímica para Bioingeniería el primer semestre de 2015, quienes trabajaron en parejas o tríos.

El instrumento (versión 2) se presentó y discutió previo al inicio de las sesiones de laboratorio. Los informes se recibieron 10 días después de la última sesión y dos docentes de la asignatura calificaron cada informe

con la rúbrica independientemente, basándose en que se ha reportado que la participación de dos evaluadores es suficiente para producir acuerdo interjueces aceptable (Jonsson y Svingby, 2007). Finalmente, se realizó una sesión de retroalimentación utilizando la rúbrica en la cual también se obtuvo la percepción del estudiante del uso del instrumento. Para ello se construyó una encuesta compuesta por 14 ítems, distribuidos en 4 dimensiones, que evaluaban el uso del instrumento por los alumnos. Estas dimensiones indagaban sobre el uso de la rúbrica previo al informe, luego de su entrega y calificación, la calidad de la construcción de la rúbrica y la utilidad de ésta. Los estudiantes debían contestar, utilizando una escala Likert de 1 a 5, donde 1 representaba total desacuerdo, 2 parcial desacuerdo, 3 ni acuerdo ni desacuerdo, 4 parcial acuerdo y 5 total acuerdo.

Finalmente, la rúbrica se modificó nuevamente considerando los comentarios de alumnos y de los docentes que evaluaron los informes. La versión definitiva de la rúbrica (versión 3) es la que se propone para ser utilizada en el futuro.

RESULTADOS

La rúbrica diseñada fue evaluada por 7 docentes expertos. El porcentaje de acuerdo interjueces para cada aspecto a evaluar se presenta en la tabla 1. Respecto de la concordancia entre los jueces, se obtuvo como resultado un Kappa de Fleiss de 0.66, lo que se califica como una adecuada concordancia entre los jueces que evaluaron la rúbrica.

Tabla 1: Evaluación de la rúbrica por pares expertos

<i>Aspectos a evaluar</i>	<i>% Acuerdo interjueces</i>
Las dimensiones agrupan todos los aspectos relevantes del tema a evaluar	100
Los criterios representan desempeños diferentes y progresivos	85.71
Las expectativas de desempeño para cada criterio son claras y medibles	71.43
El mayor criterio representa un desempeño ejemplar realista y suficiente, por ejemplo no se utiliza "destaca en"	100
El menor criterio no incluye juicios o utiliza un lenguaje ofensivo, por ejemplo "evidencia flojera"	85.71
El lenguaje utilizado es claro, simple y comprensible para el estudiante	100
Se evitan expresiones o palabras ambiguas o difíciles de definir, por ejemplo "actitud positiva", "poco", "bastante"	71.43
Los criterios se plantean positivamente y sólo se utilizan términos negativos en algunos de los criterios menores	100
¿Es la rúbrica práctica y manejable respecto a su extensión y legibilidad?	100
¿Considera que esta rúbrica es consistente con el objetivo para el cual fue construida?	100

Los jueces concordaron en su totalidad en que el instrumento incluía los aspectos relevantes al tema, con criterios que representaban desempeños suficientes y realistas, redactados desde un punto de vista positivo y con lenguaje apropiado, y que sería útil en la práctica docente. El menor acuerdo (menor a 80%) se detectó en que las expectativas de desempeño eran medibles y que la rúbrica carecía de expresiones o palabras ambiguas, por lo tanto estos aspectos fueron revisados y corregidos. Así también se realizaron modificaciones específicas en el texto sugeridas por los evaluadores y se respondió a la solicitud de explicitar detalles y/o ejemplos como notas al pie de página.

Utilizando la rúbrica corregida, se intervino el curso de bioquímica para Bioingeniería el primer semestre de 2015, presentando y discutiendo el instrumento al inicio del laboratorio. Los informes fueron evaluados por 2 docentes de forma independiente. Los puntajes de cada evaluador, diferencia porcentual entre ambos por informe y número de dimensiones que presentaron diferencias en la puntuación entre ambos evaluadores se presenta en la tabla 2.

Para estimar la concordancia de los evaluadores que utilizaron la rúbrica para calificar los informes, se utilizó el índice de correlación de r de Pearson. La correlación entre ambos evaluadores es de 0.87, lo que se interpreta como una correlación moderada-alta. Esto implica que existe una alta simetría entre los jueces al utilizar la rúbrica para calificar los informes de los estudiantes. Los puntajes fluctuaron entre 33 y 57 de un total de 64 puntos. La discrepancia entre evaluadores fue en promedio un 4.1% del puntaje total. Un solo informe presentó una discrepancia mayor a 10% (columna 4, tabla 2).

Tabla 2: Evaluación de los informes. Pje. = Puntaje obtenido de un máximo de 64 puntos, * Promedio de las diferencias 4.1, ** Número de dimensiones con diferencias de un total de 9.

Informe	Pje. Evaluador 1	Pje. Evaluador 2	Diferencia (% puntaje total)*	Dimensiones con diferencias**
1	42	43	1.6	2
2	45	40	7.8	2
3	46	46	-	-
4	46	50	6.3	3
5	51	50	1.6	1
6	56	48	12.5	3
7	35	33	3.1	1
8	57	57	-	-

Con el objetivo de confirmar que puntajes totales similares correspondan efectivamente a concordancias en las dimensiones y no simplemente a un valor total similar, en la columna 5 de la tabla 2 se presenta el número de dimensiones que presentan diferencias entre los evaluadores. Este correlaciona con el porcentaje de diferencia. Por lo tanto, las concordancias son reales, coincidiendo un bajo o nulo número de dimensiones con diferencias en informes evaluados con puntajes similares.

En la tabla 3 se presenta el número de informes con divergencia entre los dos evaluadores para cada una de las dimensiones. Se identifica que las dimensiones con mayor divergencia son introducción, resultados y apreciación global eran las dimensiones con mayor divergencia, las cuales coinciden con falencias descubiertas al momento de evaluar utilizando la rúbrica corregida posterior al juicio de expertos (versión 2).

Tabla 3: Frecuencia de divergencia por dimensión. N° = Número

Dimensión	N° Informes con divergencias entre evaluadores
Resumen 1X	2
Introducción 2X	3
Métodos 2X	-
Resultados 3X	3
Discusión 3X	2
Responsabilidad 1X	-
Formato 1X	-
Ortografía y gramática 1X	-
Apreciación global 2X	3

Consistentemente, en introducción, resultados y apreciación global se detectaron desempeños que no estaban contemplados en su totalidad en ninguno de los criterios. Esto puede haber producido que los evaluadores tomaran decisiones divergentes para optar por un criterio que no representaba fielmente el desempeño observado, el cual de haber existido, podrían haber aumentado la concordancia entre ambos jueces. Estas falencias fueron consideradas y la rúbrica se modificó en base a los desempeños observados en la experiencia de revisión. La versión final (versión 3) de la rúbrica analítica de desempeño para evaluar informes de laboratorio en formato publicación científica se presenta la tabla 4. La rúbrica posee 9 dimensiones y 4 niveles de desempeño e incluye clarificaciones sobre algunos conceptos como notas al pie de página. Por otra parte, para analizar si el utilizar la rúbrica influye en el desempeño de la tarea de confección del informe, se compararon las calificaciones del informe y de la calificación final de la asignatura completa del año 2014 y 2015, esto último con el objetivo de atribuir cualquier diferencia al desempeño en el informe a la utilización de la rúbrica y no a un desempeño superior o inferior en general en la asignatura. No se observaron diferencias significativas en los promedios de los informes (5.2 y 5.3, respectivamente) ni en los promedios de la nota final de la asignatura (5.0 y 5.1, respectivamente).

Tabla 4: Rúbrica analítica de desempeño para evaluar un informe de laboratorio formato publicación científica

¹Correctamente se refiere a mencionar el objetivo para el cual la metodología fue diseñada e incluir las etapas relevantes. ² Debe describir las metodologías y resultados de: Método de Bradford, SDS-PAGE, western blot, curva de progreso, curva de saturación, curva de dobles recíprocos, curva de pH óptimo, estudio del efecto de inhibidores enzimáticos. ³ Dedicación se refiere a evidenciar trabajo por ejemplo al seleccionar las imágenes, la bibliografía apropiada, formatear las tablas o destacar las diferentes secciones del informe; facilitando la lectura y comprensión del lector. ⁴ Despreocupación se refiere a evidenciar poco esfuerzo, el opuesto a dedicación.

Dimensión/Criterio	Muy bueno (4)	Bueno (3)	Regular (2)	Deficiente (1)
Resumen (x1)	Presenta brevemente el tema, incluyendo las metodologías y resultados principales, mencionando la relevancia del trabajo. Incluye 4 a 6 palabras clave.	Presenta el tema, incluyendo las metodologías y resultados principales, pero no menciona la relevancia del trabajo. Incluye 4 a 6 palabras clave.	Menciona la relevancia del trabajo pero omite las palabras clave o alguna metodología o resultado relevante.	No menciona la relevancia del trabajo y omite las palabras clave o alguna metodología o resultado relevante.
Introducción (x2)	Presenta el problema biológico sustentado por bibliografía, resumiendo la investigación relevante, convergiendo hacia objetivos claros y precisos.	El problema no está revisado completamente, a pesar de que sí presenta alguna fuente bibliográfica y objetivos claros y precisos.	El problema está revisado completamente, sustentado por bibliografía. Sin embargo, Los objetivos no son claros y/o precisos.	El problema no está revisado completamente. Carece de fuentes bibliográficas. Los objetivos no son claros y/o precisos.
Métodos (x2)	Describe breve, clara y correctamente ¹ todas las metodologías utilizadas ² , incluyendo citas bibliográficas cuando corresponde.	Incluye todas las metodologías utilizadas, pero la descripción de alguna es incorrecta, poco clara o extensa, o faltan citas bibliográficas.	Omite una de las metodologías utilizadas o la descripción de dos de las metodologías utilizadas son incorrectas, poco claras o extensas.	Omite dos o más de las metodologías utilizadas o la descripción de tres o más de las metodologías utilizadas son incorrectas, poco claras o extensas.
Resultados (x3)	Presenta todos los resultados ² , organizados en tablas o gráficos cuando corresponde, los cuales poseen variables y unidades en todos los casos.	Presenta todos los resultados, organizados en tablas o gráfico cuando corresponde, sin embargo, uno o más carece(n) de la variable o de la unidad.	Presenta todos los resultados, organizados en tablas o gráfico cuando corresponde, sin embargo, comete errores conceptuales al describir alguno(s) de ellos. <i>*Obtendrá este criterio aunque los resultados posean variables o unidades.</i>	Omite el resultado de alguno de los experimentos. <i>*Obtendrá este criterio aunque el resto de los resultados posean variables o unidades o los describa correctamente.</i>

Tabla 4 (continuación)

Discusión de los resultados (x3)	Compara los resultados con la bibliografía, discutiendo sobre su calidad considerando las limitaciones de las metodologías utilizadas. Finaliza con una conclusión clara y breve basada en resultados comprobados.	Compara los resultados con la bibliografía, discutiendo sobre su calidad considerando las limitaciones de las metodologías utilizadas. Carece de una conclusión final.	La discusión está incompleta, ya sea por falta de base bibliográfica o de análisis de los resultados. Sin embargo, finaliza con una conclusión clara y breve basada en resultados comprobados.	La discusión está incompleta, ya sea por falta de base bibliográfica o de análisis de los resultados. Carece de una conclusión final.
Responsabilidad (x1)	Entrega en la fecha estipulada y el trabajo está prolijamente presentado.	Entrega en la fecha estipulada, pero falta prolijidad, por ejemplo está arrugado o sucio.	El trabajo esta prolijamente presentado, peor entrega fuera de plazo.	Entrega fuera de plazo y falta prolijidad.
Formato (x1)	El informe incluye todas las secciones, respetando formato y longitud de acuerdo a las instrucciones.	El informe incluye todas las secciones, pero no respeta formato y/o longitud especificada.	El informe carece de alguna de las secciones. <i>*Obtendrá este criterio aunque el resto de las secciones respeten el formato y longitud.</i>	El informe carece de dos o más secciones. <i>*Obtendrá este criterio aunque el resto de las secciones respeten el formato y longitud.</i>
Ortografía y gramática (x1)	El texto es coherente, identificándose ideas principales y secundarias. Sigue las normas gramaticales, de ortografía y puntuación del idioma español.	El texto es coherente, identificándose ideas principales y secundarias. Presenta 1 o 4 errores de ortografía o puntuación del idioma español, por página.	El texto es coherente, identificándose ideas principales y secundarias. Presenta 5 o más errores de ortografía o puntuación del idioma español, por página.	Las ideas principales no se diferencian de las secundarias. Presenta 5 o más errores de ortografía o puntuación del idioma español, por página.
Apreciación global (x2)	El trabajo denota dedicación ³ , comprensión del tema y del trabajo científico.	El trabajo denota comprensión del tema y del trabajo científico, pero se aprecia confusión y/o despreocupación ⁴ en algunos aspectos.	El trabajo denota comprensión del tema y del trabajo científico, pero falta análisis crítico de los resultados	Hay más aspectos que sugieren confusión que comprensión.

La percepción del estudiante de la experiencia de uso de la rúbrica en bioquímica se estimó mediante una encuesta con una escala Likert de 1 a 5, que incluía aspectos previos y posteriores a la entrega del informe, además de aspectos relacionados con la construcción y utilidad del instrumento. El promedio de los puntajes del total de los estudiantes en los diferentes ítems, y en las dimensiones que estos evaluaban, se presentan en la tabla 5.

Tabla 5: Percepción del estudiante. *Corresponde al promedio de los puntajes asignados con una escala Likert de 1 (total desacuerdo) a 5 (total acuerdo)

<i>Dimensión Evaluada</i>	<i>Promedio*</i>
PREVIO A LA ENTREGA DEL INFORME	
Revisé la rúbrica detalladamente antes de confeccionar el informe.	4
La rúbrica me guio en la confección del informe.	3.93
Promedio	3.97
LUEGO DE LA ENTREGA Y CALIFICACIÓN	
La rúbrica me retroalimentó sobre mi desempeño.	4
Concuerdo con la calificación obtenida con la rúbrica.	3.71
Promedio	3.86
CONSTRUCCIÓN DEL INSTRUMENTO	
Los criterios para cada desempeño están claramente definidos.	3.93
Los criterios se diferencian entre un nivel desempeño y otro.	3.86
Todos los aspectos importantes están considerados en la rúbrica.	3.93
El lenguaje utilizado es claro, simple y comprensible.	4.36
Me parece de una longitud adecuada y no me distraigo al leerla	3.5
Promedio	3.92
UTILIDAD DEL INSTRUMENTO	
La rúbrica me parece adecuada para evaluar el informe de laboratorio.	4.01
La rúbrica me ayuda a completar la tarea asignada.	3.71
Me parece que utilizarla aumenta la objetividad en la corrección.	4.36
Me gustó utilizar la rúbrica.	3.86
Me gustaría utilizar más frecuentemente rúbricas para evaluar esta u otras tareas en esta u otras asignaturas.	4.5
Promedio	4.01

Los promedios por dimensión fluctúan entre 3.86 y 4.01, con promedios en todos los aspectos evaluados iguales o mayores a 3.5, por lo que la evaluación de todos los ítems es mayoritariamente positiva. Los ítems con mayor puntaje corresponden a que les gustaría utilizar rúbricas más frecuentemente en esta u otras asignaturas, que la asocian con un aumento de objetividad en la corrección y a que el lenguaje utilizado es claro, simple y comprensible. Cabe señalar que declaran revisar el instrumento antes de confeccionar el informe, y que el instrumento los retroalimentó sobre su desempeño. El menor promedio se detectó para el ítem que se refiere a la longitud y distracción al leerla.

DISCUSIÓN

Actualmente es una práctica creciente el explicitar las descripciones de estándares de calidad, por ejemplo a través de rúbricas, basándose en la creencia de que se generarán niveles apropiados de consistencia entre evaluadores y comparabilidad en el logro de los estudiantes (Sadler, 2014), además de las múltiples ventajas previamente mencionadas para este instrumento. En este contexto, este manuscrito describe una experiencia en el aula para actualizar la evaluación del producto del laboratorio de un curso de bioquímica, un informe global en formato publicación científica, mediante la creación, validación por juicio de expertos y uso de una rúbrica analítica de desempeño.

Reddy y Andrade (2010) reportan que los docentes se resisten a elaborar y utilizar rúbricas, lo cual es apoyado por Allen y Tanner (2006), que mencionan lo trabajoso que es su elaboración, la descripción de modo que esta experiencia podría contribuir a que otros diseñen o perfeccionen instrumentos creando versiones que sean validadas por pares expertos y perfeccionadas después de utilizarlas.

Se presenta la rúbrica (versión 3) en la tabla 4, la cual puede ser útil para otros docentes dedicados a la enseñanza de las ciencias. Esta versión final requirió de dos versiones previas para mejorar la claridad de algunos criterios, evidenciando que el lenguaje es lo más desafiante, en el sentido de que sea claro y apropiado (Reddy y Andrade, 2010), requiriendo explicitar algunas definiciones de conceptos y en algunos casos ejemplos. Por otra parte, de la experiencia se desprende que se requiere pilotear el uso para verificar que todos los desempeños posibles estén considerados y detectar otras necesidades de mejora. Los resultados del uso del instrumento en la cohorte 2015 sugieren que el instrumento es confiable, presentando un consenso entre jueces, mayor a lo reportado por Jonsson y Svingby (2007), quienes declaran que el consenso es de 70%. Consecuentemente, los estudiantes perciben que aumenta la objetividad en la corrección, a pesar de que el acuerdo con la calificación obtenida no es absoluto. El uso de ejemplos puede haber contribuido a la confiabilidad.

Los más indicados para evaluar si las rúbricas promueven el aprendizaje son los estudiantes y profesores que la utilizan, de modo que es importante que la percepción sea parte del estudio. (Jonsson y Svingby, 2007). Pensando en la rúbrica como un instrumento que retroalimenta al estudiante, en lo cual hay neutralidad o acuerdo en esta experiencia, Sadler (2010) enfatiza en que hay que centrarse en las interpretaciones de la retroalimentación de los estudiantes, en cómo éste percibe la retroalimentación en lugar de cómo el profesor lo intenciona, lo que nuevamente otorga relevancia a la opinión de los estudiantes. En esta experiencia, los estudiantes declaran utilizar el instrumento y lo consideran adecuado para lo que fue diseñado, consideran el lenguaje apropiado, que se incluye los aspectos importantes, con criterios claros. A pesar de ello, no hay acuerdo en que los guía en la confección, declarando que es extensa y los distrae. Hay neutralidad y parcial acuerdo en que les agrada utilizarla, sin embargo, hay acuerdo parcial y total en que les gustaría utilizarla en otra asignatura.

Por otra parte, el incorporar el instrumento no tuvo efecto en el rendimiento de la asignatura. Sadler (2014) ha reportado que instrucciones y retroalimentación sobre las fortalezas y debilidades del trabajo en algunos estudiantes tiene nulo o bajo impacto, planteando que se debe a dificultades para comprender el significado de los criterios y detectar aspectos claves que requieren atención, demandando enriquecer la propuesta más allá de presentar los criterios en la rúbrica, para lograr la comprensión de los estándares de calidad. La rúbrica también puede ser utilizada por el profesor para identificar, caracterizar y clasificar dificultades en la tarea asignada, para luego enfocarse en esas dificultades para promover el aprendizaje. Es así como Dasgupta et al. (2014) identificaron las dificultades de los estudiantes en lograr los criterios para el diseño experimental. En esta experiencia, se identificó que los estudiantes tienen dificultades para analizar datos, para concluir desde datos experimentales, ya que la discusión fue la dimensión con menor puntaje (dato no mostrado).

Wilson et al. (2016) han reportado que los procesos para evaluar están desalineados con el aprendizaje deseado. Al incluir rúbricas y actividades que ayuden a comprender criterios y estándares de calidad, como se llevó a cabo en esta intervención, se está avanzando en este respecto. Se reporta confiabilidad al evaluar, se detectaron aspectos del desempeño de los estudiantes que necesitan mejorar y los estudiantes valoran positivamente el instrumento. Sin embargo, no hay efecto en las calificaciones evidenciando que se requiere robustecer el diseño de la intervención. En este contexto, con el desafío de enfocarse en el proceso y no principalmente en los productos de las actividades de enseñanza-aprendizaje y otorgarle un rol fundamental a la evaluación y retroalimentación centrada en el estudiante en el aula, se propone incorporar las siguientes acciones al utilizar el instrumento:

1. *Incluir una sesión de definición de estándares previo a presentar el instrumento utilizando discusión grupal de informes "ejemplo".* Sadler (2014) reporta que el analizar ejemplos contribuye a utilizar un instrumento de evaluación efectivamente y tener conocimiento sobre los estándares y criterios. Así mismo, O'Donovan, Price, y Rust (2008), plantean que el uso de ejemplos e involucrar a los estudiantes puede mejorar su desempeño significativamente. Consecuentemente, Tai et al. (2018) reportan que puede contribuir al desarrollo del juicio evaluativo, que se refiere a la capacidad de tomar decisiones sobre la calidad de su trabajo y el de otros, lo cual se ha planteado como una meta de la educación superior.

2. *Incluir una sesión para presentar gráficos y tablas obtenidas para discusión grupal para análisis crítico de datos y resultados científicos al finalizar las sesiones de laboratorio previo a la elaboración del informe.* Esta acción responde a las dificultades de los estudiantes para analizar datos, que se detectaron en esta experiencia, a modo de procurar que logren la habilidad.

3. *Pre-entrega para evaluación por pares.* Al involucrar al estudiante a través de la evaluación por pares se puede propiciar la autodirección y ayudarlos a entender como ellos y otros aprenden (Allen y Tanner, 2006). Li y Gao (2016) reportan a la evaluación por pares como una herramienta efectiva en el desempeño de aprendizaje por proyecto, los estudiantes pueden involucrarse en su aprendizaje preparándolos para autoevaluación y reflexión como pensadores críticos para la resolución de problemas, promoviendo que comprendan los criterios de evaluación y en consecuencia, mejorando la calidad de desempeño.

4. *Incluir una sesión de retroalimentación individual final para proveer acciones concretas para mejorar.* Se ha reportado que la rúbrica es una herramienta motivacional poderosa cuando los profesores proveer retroalimentación constructiva individual (Wu, Heng, y Wang, 2015).

CONCLUSIONES

En este trabajo se describe el diseño y uso de un instrumento para evaluar una tarea de desempeño que tiene por objetivo acercar el aula al trabajo científico, cuya versión final se comparte. A partir del resultado y reflexión se generó una propuesta que incorpora el valor de la evaluación por pares y retroalimentación para el logro de aprendizaje, procurando que la experiencia se centre en el proceso y sea útil para los estudiantes. Esta propuesta cumple con las 3 características claves de la evaluación para el aprendizaje (assessment for learning): 1) la importancia del diseño de la evaluación para promover y sostener el aprendizaje de los estudiantes; 2) Proveer de retroalimentación que les permita mejorar su trabajo; 3) Claridad en los estándares para entender la diferencia entre los desempeños obtenidos y esperados (Tan, 2013). La experiencia ilustra el requerimiento y desarrollo de una práctica docente de mejoramiento y actualización continua acorde con los objetivos y logros de los estudiantes.

REFERENCIAS

- Alberts, B., Considering Science Education (Editorial), *Science*, 319(21), 1189 (2008)
- Allen, D. y K. Tanner, Rubrics: Tools for making learners goals and evaluation criteria for both teachers and learners, doi: 10.1187/cbe.06-06-0168, *Cell Biology Education*, 5, 197-203 (2006)
- Barr, R.B. y J. Tagg, From teaching to learning - a new paradigm for undergraduate education, doi: 10.12691/education-3-9-6, *Change*, 27(6), 13-25 (1995)
- Biggs, J., What the student does: teaching for enhanced learning, doi:10.1080/0729436990180105, *Higher Education Research & Development*, 18(1), 57- 71 (1999)
- Boud, D. y N. Falchikov, *Rethinking Assessment in Higher Education, Learning for the longer term*, Routledge, New York, Estados Unidos (2007)
- Boud, D. y E. Molloy, Rethinking models of feedback for learning: the challenge of design, doi: 10.1080/02602938.2012.691462, *Assessment & Evaluation in Higher Education*, 38(6), 698-712 (2013)
- Bess, M.T., J. Sancho y otros cuatro autores, Purification of colored photosynthetic proteins for understanding Isolation principles, doi:10.1002/bmb.2003.494031020202, *Biochemistry and Molecular Biology Education*, 31(2), 119-122 (2003)
- Bloxham, S., Central challenges in transforming assessment at departmental and institutional level, Keynote address at the Assessment in Higher Education (AHE) seminar day, Manchester (2016)
- Cargill, M. y P. O'Connor, *Writing Scientifics Research Articles: Strategy and Steps*. Wiley-Blackwell, Chichester, England (2011)
- Dasgupta, A.P., T.R. Anderson y N. Palaez, Development and validation of a rubric for diagnosing students' experimental design knowledge and difficulties, doi: 10.1187/cbe.13-09-0192, *Cell Biology Education*, 13, 265-284 (2014)
- Domenech, J., Secuencias de apertura experimental y escritura de artículos en el laboratorio: un itinerario de mejora de los trabajos prácticos en el laboratorio, *Enseñanza de las Ciencias*, 31(3), 249-262 (2013)
- Earl, L., *Assessment as Learning: Using Classroom Assessment to Maximize Student Learning*, Corwin Press Press, California, Estados Unidos (2013)
- Figuroa, A.E., M.C. Gilio y V.F. Gutiérrez, La función docente en la universidad, *Revista Electrónica de Investigación Educativa*, Número Especial, 1-14 (2008)
- González, C., Perspectivas y desafíos de la docencia en la educación superior, en: *La Educación Superior en Chile, Transformación, Desarrollo y Crisis*, de: A. Bernasconi, Ediciones Universidad Católica de Chile, Santiago, Chile (2015)
- Harden, R.M y J. Crosby, AMEE Guide No 20: The good teacher is more than a lecturer- the twelve roles of the teacher, doi:10.1080/014215900409429, *Medical Teacher*, 22(4), 334-347 (2000)
- Ifeanyichukwu, H., Assessing Laboratory Skills Performance in Undergraduate Biology Students, doi:10.5901/ajis.2016.v5n3p113, *Academic Journal of Interdisciplinary Studies*, 5(3), 113-122 (2016)
- Jonsson, A. y G. Svingby, The use of scoring rubrics: Reliability, validity and educational consequences, doi: 10.1016/j.edurev.2007.05.002, *Educational Research Review*, 2, 130-144 (2007)
- Martínez-Rizo, F. y A. Mercado, Estudios sobre prácticas de evaluación en el aula: revisión de la literatura, *Revista Electrónica de Investigación Educativa*, 17(1), 17-32 (2015)
- Nicol, D. y D. Macfarlane-Dick, Formative assessment and self-regulated learning: A model and seven principles of good feedback practice, doi: 10.1080/03075070600572090, *Studies in Higher Education*, 31(2), 199-218 (2006)
- O' Donovan, B., M. Price y C. Rust, Developing student understanding of assessment standards: a nested hierarchy of approaches, doi:10.1080/13562510801923344, *Teaching in Higher Education*, 13(2), 205-217 (2008)

- Ohaja, M., M. Dunlea y K. Muldoon, Group marking and peer assessment during a group poster presentation: The experiences and views of midwifery students, doi:10.1016/j.nepr.2012.11.005, *Nurse Education in Practice*, 13, 466-470 (2013)
- Pales, J., Planificar un currículum o un programa formativo, *Educación Médica*, 9(2), 59-65 (2006)
- Palm, T., Performance assessment and authentic assessment: a conceptual analysis of the literature, doi: 10.1080/02602938.2015.1085954, *Practical Assessment, Research and Evaluation*, 13(4), 1-11 (2008)
- Prins, F.J., R. de Kleijn y J. van Tartwijk, Students' use of a rubric for research theses, doi: 10.1080/02602938.2015.1085954, *Assessment & Evaluation in Higher Education*, 42(1) 128-150 (2017)
- Raposo, M. y E. Martínez, La rúbrica en la enseñanza universitaria: Un recurso para la tutoría de grupos de estudiantes, doi:10.4067/S0718-50062011000400004, *Formación Universitaria*, 4(4), 19-28 (2011)
- Reddy, Y.M. y H. Andrade, A review of rubric use in higher education, doi:10.1080/02602930902862859, *Assessment & Evaluation in Higher Education*, 35(4), 435-448 (2010)
- Ribeiro, R. y M. Assunção, Conceptions and Practices of Assessment in Higher Education: A Study of Portuguese University Teachers, doi: 10.15366/riee2016.9.1.001, *Revista Iberoamericana de Evaluación Educativa*, 9(1), 9-29 (2016)
- Rudner, L.M. y C. Boston, Performance Assessment, *ERIC Review*, 3(1), 2- 12 (1994)
- Stanford School Redesign Network What is performance – based assessment? (2008)
- Sadler, D.R., Beyond feedback: developing student capability in complex appraisal, doi: 10.1080/02602930903541015, *Assessment & Evaluation in Higher Education*, 35(5), 535-550 (2010)
- Sadler, D.R., The futility of attempting to codify academic achievement standards, doi: 10.1007/s10734-013-9649-1, *Higher Education*, 67, 273-288 (2014)
- Şanlı, Ö., S. Erdem y T. Tefik, How to write a discussion section? doi:10.5152/tud.2013.049, *Turkish Journal of Urology*, 39(1), 20-4 (2013)
- Sambel, K., Assessment and feedback in higher education: considerable room for improvement? *Student Engagement in Higher Education Journal*, 1(1), 1-14 (2019)
- Schell, J. y J. Porter, Applying the Science of Learning to Classroom Teaching: The Critical Importance of Aligning Learning with Testing, doi: 10.1111/1541-4329.12141, *Food Science Education*, 17, 36-41 (2018)
- Stiggins, R., Assessment Crisis: The absence of assessment for learning, doi:10.1155/2013/640609, *Phi Delta Kappan*, 83(10), 758-65 (2002)
- Tai, J., R. Ajjawi, D. Boud, P. Dawson y E. Panadero, Developing evaluative judgement: enabling students to make decisions about the quality of work, doi: 10.1007/s10734-017-0220-3, *Higher Education*, 76(3), 467-481 (2018)
- Tan, K., A framework for Assessment for Learning: Implications for feedback practices within and beyond the gap, doi: 10.1155/2013/640609, *International Scholarly Research Notices Education*, Article ID 640609, 6p. (2013)
- Vanaki, Z. y R. Memarian, Professional Ethics: beyond the clinical competency, doi: 10.1016/j.profnurs.2009.01.009, *Journal Professional Nursing*, 25, 285-291 (2009)
- Villarroel, V. y D. Bruna, Reflexiones en torno a las competencias genéricas en educación superior: Un desafío pendiente, *Psicoperspectivas*, 13(1), 23-34 (2014)
- Villarroel, V., D. Bruna y otros tres autores, Implementación de la Evaluación Auténtica en Educación Superior, in *Innovando en Educación Superior: Experiencias Clave en Latinoamérica y el Caribe 2016-2017*, Vol. 2: Metodologías activas de enseñanza y aprendizaje, por O. Jerez y C. Silva, 1ª Ed., 249-258, Centro de Enseñanza y Aprendizaje Universidad de Chile y LASPAU, afiliado con Universidad de Harvard, Santiago, Chile (2017)
- Villarroel, V., C. García y otros tres autores, Aprender del error es un acierto, Las dificultades que enfrentan los estudiantes chilenos en la Prueba PISA, doi:10.4067/S0718-07052015000100017, *Rev. Estudios Pedagógicos*, 41(1), 293-310 (2015)
- Villarroel, V., S. Bloxham y otros tres autores, Authentic assessment: creating a blueprint for course design, doi: 10.1080/02602938.2017.14123969, *Assessment & Evaluation in higher Education*, 43(5), 840-854 (2018)
- Waldrop, M., The sciences of teaching science, doi: 10.1038/523272a, *Nature*, 523, 272-274 (2015)
- Wilson, A., S. Howitt y D. Higgins, A fundamental misalignment: intended learning and assessment practices in undergraduate science research projects, doi:10.1080/02602938.2015.1048505, *Assessment & Evaluation in Higher Education*, 41(6), 869-884 (2016)
- Windschitl, M., Inquiry Projects in Science Teacher Education: What Can Investigative Experiences Reveal about Teacher Thinking and Eventual Classroom Practice? doi: 10.1002/sce.10044, *Science Education*, 87(1), 112 –143 (2003)
- Wu, X., M. Heng y W. Wang, Nursing students' experiences with the use of authentic assessment rubric and case approach in the clinical laboratories, doi: 10.1016/j.nedt.2014.12.009, *Nurse Education Today*, 35, 549-555 (2015)