

Dictámenes en Revistas Científicas: lo que Necesitan los Editores y Autores, lo que Hacen los Evaluadores

Journal Referee Reports: What Editors and Authors Need, What Referees Do

Ditames em Revistas Científicas: o que os Editores e Autores Precisam, o que os Avaliadores Fazem

TATIANA PLATA-CAVIEDES

Universidad de los Andes, Bogotá, Colombia

OSCAR ALEJANDRO CÓRDOBA-SALGADO

Horizontes ABA Terapia Integral, Bogotá, Colombia

PIOTR TRZESNIAK

Universidade Federal de Itajubá, Itajubá, Brasil

Resumen

Con el objetivo de conocer si los pares evaluadores que participaron en la *Revista Colombiana de Psicología*, en los años 2010 y 2011, proporcionaron al editor los recursos para tomar una decisión fundamentada sobre la publicación de un artículo, así como herramientas al autor para mejorar su manuscrito, se realizó un análisis de contenido a 96 conceptos. Se encontró que los evaluadores se centraron más en aspectos relacionados con el mérito científico del artículo e hicieron comentarios precisos; sin embargo, solo la mitad de ellos realizaron comentarios constructivos. Se discute la relevancia de la responsabilidad de los pares evaluadores de mejorar la calidad de los artículos, además de decidir sobre su publicación.

Palabras clave: revisión por pares, publicación científica, procesos editoriales.

Abstract

In order to determine whether or not the peer reviewers who worked with the *Revista Colombiana de Psicología* in 2010 and 2011 provided the editor with the resources necessary to make an informed decision regarding publication decisions and the authors with the tools needed to improve their manuscripts, a content analysis of 96 concepts was conducted. Findings showed that referees concentrated more on aspects related to the scientific value of the article, and provided precise comments in most cases. However, only half of the reviewers provided constructive feedback. The article discusses the reviewers' responsibility in improving the quality of the articles, in addition to deciding on their publication.

Keywords: peer review, scientific publication, editorial processes.

Resumo

Com o objetivo de conhecer se os pares avaliadores que participaram na *Revista Colombiana de Psicología*, nos anos 2010 e 2011, proporcionaram ao editor os recursos para tomar uma decisão fundamentada sobre a publicação de um artigo, assim como ferramentas para o autor melhorar seu manuscrito, realizou-se uma análise de conteúdo a 96 conceitos. Encontrou-se que os avaliadores se centraram mais em aspectos relacionados com o mérito científico do artigo e fizeram comentários precisos; contudo, só a metade deles realizaram comentários construtivos. Discute-se a relevância da responsabilidade dos pares avaliadores de melhorar a qualidade dos artigos, além de decidir sobre sua publicação.

Palavras-chave: revisão por pares, publicação científica, processos editoriais.

La correspondencia en relación con este artículo puede dirigirse a cualquiera de los autores. Tatiana Plata-Caviedes, e-mail: tatianaplata@gmail.com, Oscar Alejandro Córdoba-Salgado, e-mail: oacordobas@gmail.com y a Piotr Trzesniak, e-mail: piotrze@gmail.com. Departamento de Física e Química, Universidade Federal de Itajubá, Av. BPS 1303, Campus Universitario 37500-903, Itajubá.

ARTÍCULO DE INVESTIGACIÓN CIENTÍFICA
RECIBIDO: 1 DE FEBRERO DE 2012 - ACEPTADO: 16 DE MARZO DE 2012

EN LA actualidad, existen aproximadamente 500 revistas científicas de psicología en Iberoamérica, número que está en constante ascenso (Buela-Casal & López, 2005). En Colombia, para el año 2008, se identificaron un total de 24 publicaciones vigentes y, al igual que lo que ocurre a nivel iberoamericano, existe un aumento significativo del número de revistas y artículos científicos publicados. En esta línea, en el periodo que va desde el año 1991 a 1997, se lanzaron en Colombia cinco nuevas revistas y se publicaron 595 artículos (62% eran artículos científicos); mientras que, en el periodo que va de 1998 a 2008, se crearon 22 revistas nuevas y se publicaron 3,052 artículos, de los cuales el 80% eran artículos científicos (Guerrero & Jaraba, s.f.).

Ahora bien, a pesar del número creciente de artículos en el área, son muy pocas las revistas en Latinoamérica que se encuentran indexadas en índices internacionales de gran prestigio como SCOPUS o ISI; solo cuatro revistas latinoamericanas de psicología se encuentran indexadas en ambas bases y 19 están en SCOPUS. Además, no todos los países tienen revistas de psicología en estas bases, únicamente Brasil, México, Argentina, Colombia y Chile (Vera-Villarroel, López-López, Lillo & Silva, 2011).

El número de revistas aceptadas por SCOPUS en Latinoamérica va en aumento, así como la visibilidad internacional de algunas revistas de la región. Ejemplos claros de esto son la *Revista Latinoamericana de Psicología* (Colombia), *Universitas Psychologica* (Colombia) y la *Revista Latinoamericana de Psicopatología Fundamental* (Brasil), según lo indica el índice de internacionalidad de Zych y Buela-Casal (2010). Pero todavía es largo el camino que deben seguir las revistas de la región para lograr mejores resultados y un mejor posicionamiento a nivel internacional.

Una forma prioritaria de lograr estos resultados es centrarse en el mejoramiento de los procesos editoriales, es decir, de los procesos que permiten publicar artículos científicos de gran

calidad, que sean relevantes para la comunidad académica y, por lo tanto, generen un impacto en la misma. Más allá de las estrategias políticas o de la infraestructura que puedan o no ser necesarias para lograr la internacionalización de una revista, el mejoramiento de los procesos editoriales es la estrategia más benéfica para la comunidad académica y la construcción de conocimiento. Es por eso que este texto se centrará en el proceso editorial de un artículo científico, específicamente, en el proceso de evaluación por pares, debido a su papel fundamental en las publicaciones de calidad científica.

Originalmente, se puede dividir el proceso de evaluación de un artículo científico en dos momentos, la evaluación previa y la evaluación por pares. En el primero, el editor o editor adjunto (Trzesniak, 2009) realiza una revisión inicial del manuscrito, con el objetivo de verificar si el artículo se adecúa o no al área temática de la revista, o si el artículo tiene o no una calidad muy pobre (Rowland, 2002). Este primer momento es primordial dentro del proceso, debido a que la evaluación por pares es costosa y no es rentable utilizarla si desde un inicio se detecta que el manuscrito no tiene las características mínimas de un artículo científico; por ejemplo, que no haya una justificación de la pregunta de investigación, que no se describa de manera completa la metodología empleada, que los resultados no sean claros, que no se discutan los resultados, o que las conclusiones no respondan a la pregunta de investigación. Para realizar esta evaluación, no es necesario que el evaluador conozca a profundidad la temática que se está abordando, sin embargo, es fundamental que tenga un conocimiento especializado en investigación y sobre las características fundamentales de un artículo científico.

El segundo momento dentro de la evaluación de un artículo es la evaluación por pares. Si el manuscrito cumple con los criterios mínimos de un artículo científico, este se debe evaluar en detalle, con el objetivo de determinar si, más allá

de esos criterios, el documento tiene mérito científico, es decir, si es suficientemente relevante para ser publicado. Esta labor ya no la puede hacer el editor, por dos razones básicas: (a) dependiendo de la revista, su área y su reputación, entre otros, un editor puede recibir desde decenas hasta cientos o incluso miles de artículos al año; y (b) un editor puede recibir artículos de diferentes subtemáticas, algunas veces más especializadas que otras. Estos dos factores dificultan el que un editor pueda, primero, leer cada uno de los artículos con el nivel de detalle que le permitiría publicar únicamente artículos científicos de calidad y, segundo, conocer con suficiente detalle todas las temáticas que se abordan en cada manuscrito que es sometido a evaluación. Es por esto que el editor debe apoyar su labor en los pares evaluadores.

En este punto, el editor debe enviarle a mínimo dos académicos o jurados, expertos en la temática del artículo, el manuscrito a ser evaluado. Los jurados deben realizar una evaluación del manuscrito y escribir un dictamen evaluativo en el que describen sus apreciaciones, además de comunicar al editor si el artículo debe ser (a) publicado en su estado actual (puede requerir solo algunos cambios de forma), (b) publicado después de realizar correcciones mayores (cuando hay un trabajo empírico relevante y bien realizado, pero el informe falla en presentarlo de una manera adecuada) o (c) no publicado (cuando no hay un mérito científico en el manuscrito). La decisión más frecuente es la segunda (Rowland, 2002), en la que los pares evaluadores le informan a los editores los cambios que los autores deben realizar para que su manuscrito sea publicado.

Es el editor, no el evaluador, quien finalmente decide si el artículo en cuestión se publica o no. Por esto, el dictamen de un evaluador debe permitir que el editor se informe de las características del artículo y pueda tomar una decisión informada del valor y la calidad del mismo. El evaluador debe, en términos generales, apoyar la labor del editor para que solo los artículos de

relevancia científica sean publicados; así, debe revisar los siguientes aspectos de un artículo (Benos, Kirk & Hall, 2003): (a) relevancia de la pregunta de investigación, (b) novedad y aporte del estudio, (c) debilidades y fortalezas de la metodología e interpretación de resultados, (d) estilo y presentación de figuras y tablas, y (e) cuestiones éticas. Adicionalmente, debe revisar que se discutan los resultados obtenidos y que el artículo en su totalidad sea coherente con la pregunta de investigación. Se espera que el par evaluador se centre en los aspectos que se relacionan directamente con el mérito científico del estudio que se presenta, aunque puede mencionar aspectos de tipo formal (e. g. redacción, seguimiento de normas APA, etc.) o editorial (e. g. que el manuscrito tenga las partes que componen un artículo científico). El evaluador debe enfocar sus apreciaciones en los primeros elementos, que son aquellos que solo él como experto en el área puede evaluar. Los otros dos aspectos pueden ser evaluados por un editor, asistente editorial y, en algunos aspectos, por un corrector de estilo.

Además de revisar los anteriores aspectos, es importante que, cuando redacte el dictamen, el evaluador le dé los elementos necesarios al editor para que este pueda entender sus apreciaciones y pueda sopesar dichas apreciaciones al decidir si publica o no el manuscrito (Benos et al., 2003). Para que un evaluador logre esto último, es importante que redacte en su dictamen los argumentos de cada valoración que realice y que aterrice las afirmaciones que haga sobre el contenido específico del artículo, por ejemplo, describiendo el contenido que el autor desarrolla en su manuscrito o dando ejemplos que demuestren su punto. Esto le permitirá al editor entender sin ambigüedades las apreciaciones hechas por el evaluador.

Sin embargo, el dictamen que un evaluador redacta es una herramienta que no solo será de utilidad para el editor, sino también para el autor. Como se mencionó anteriormente, un editor

puede tomar una de tres opciones principales respecto de un artículo: (a) no publicarlo, (b) publicarlo con cambios y (c) publicarlo en su estado actual. Si el editor escoge la opción a, el autor estará altamente motivado por conocer las razones por las cuales su artículo no fue aceptado; por otro lado, si es aceptado con cambios, el dictamen del evaluador será la herramienta principal para que el autor realice los cambios pertinentes y, en gran parte, dependerá de la calidad del dictamen el que el autor sea exitoso en mejorar su artículo para que pueda ser aceptado y publicado.

Es así que un dictamen, aparte de centrarse en el mérito científico del artículo y de tener valoraciones que sean claras (a través de la argumentación de cada una de ellas y de su descripción en términos concretos), debe ser constructivo. Esto significa que, para que un dictamen sea de utilidad para el autor, no debe limitarse a detectar y describir los problemas y fortalezas de un artículo, sino que debe dar sugerencias concretas acerca de los cambios a través de los cuales se puede resolver cada problema o mejorar el manuscrito en cuestión.

En resumen, el evaluador deberá actuar entonces, en palabras de Benos et al. (2003), como un abogado del autor y del editor, con el fin de facilitar que ambos cumplan con su respectivo objetivo, que será, para el caso del autor-investigador, publicar un artículo correctamente elaborado y de valor científico, y para el caso del editor, publicar únicamente los artículos que cumplan con los estándares científicos y que sean de valor para la comunidad académica.

Ahora bien, el proceso de evaluación por pares es altamente valorado por la comunidad académica. En un estudio realizado por The Association of Learned and Professional Society Publishers, un 70% de los autores consultados consideraron estar satisfechos o muy satisfechos con el proceso de evaluación por pares (Rowland, 2002). Sin embargo, son varios los estudios y artículos en el área que muestran varias dificultades o problemáticas que tiene este proceso.

Las que siguen son algunas de las problemáticas que se han detectado en el proceso de evaluación por pares: no se detectan problemas metodológicos graves, los evaluadores se sesgan por el estatus académico del autor, subestiman estudios que no obtienen resultados significativos o que son réplicas, o incluso aquellos que reportan hallazgos o teorías innovadoras (Campanario, 2002). Schroter et al. (2008) encontraron una tasa muy baja de detección de errores colocados deliberadamente en un manuscrito; los errores menos detectados fueron la falta de concordancia entre los datos y los resultados descritos, el control de variables temporales y una justificación precaria del estudio.

Este tipo de errores lleva a cuestionar los elementos en los que se están centrando los evaluadores al revisar un manuscrito y a preguntar por qué se dejan de detectar elementos de gran relevancia. Bornmann, Nast y Daniel (2008), en un meta-análisis realizado con 46 estudios sobre los criterios que los editores y evaluadores tienen para evaluar un manuscrito, encontraron que los aspectos en los que los evaluadores más se centran no son necesariamente los que los editores consideran de mayor relevancia. Por ejemplo, encontraron que los evaluadores se centran, primordialmente, en la relevancia de la contribución y la presentación del manuscrito; sin embargo, al consultar tanto a editores como evaluadores acerca de cuáles consideraban que deberían ser los criterios de mayor relevancia para decidir si aceptar o rechazar un manuscrito, los más relevantes fueron la discusión de resultados (que sea objetiva, clara y se desarrolle a profundidad) y la teoría (que el estudio aporte elementos novedosos a una teoría o que la teoría manejada sea de interés). Además, los criterios en los que los evaluadores más se centran en su práctica (relevancia de la contribución y presentación), resultaron en un nivel intermedio de relevancia. Los criterios menos relevantes fueron los criterios éticos y la afiliación y reputación del autor, aun cuando, como ya se mencionó, los

evaluadores sí se sesgan en algunas oportunidades por la reputación del autor.

Esto lleva a plantear la posibilidad de que los evaluadores pueden fallar en cumplir a cabalidad su rol cuando se centran en elementos que no son los más relevantes para el editor y el autor. Como se describió anteriormente, la labor del evaluador es ayudar a que ambos actores puedan cumplir con sus respectivos roles (que para el autor será redactar un buen artículo y para el editor publicar únicamente los artículos de calidad científica). En este sentido, el evaluador deberá centrar su evaluación en los elementos que permitan tanto al editor como al autor cumplir con sus objetivos.

Retomando lo que ya se ha dicho, se espera que el evaluador se centre en los aspectos que solo él, como experto en el área, puede detectar (fundamentación del estudio, pertinencia metodológica, relevancia del estudio y pertinencia de resultados, discusión y conclusiones); además, que dé los elementos necesarios para que el editor entienda y evalúe las apreciaciones que hace (debe argumentar y dar ejemplos concretos de cada una). Finalmente, para que el dictamen le permita al autor cumplir con su objetivo, el evaluador debe, no solo detectar los problemas del manuscrito, sino ser constructivo y dar sugerencias concretas sobre las formas en que el manuscrito puede ser mejorado.

Por todo lo anterior, con el objetivo de conocer si la forma en que los evaluadores redactaron sus dictámenes cumple con las necesidades de autores y editores, además de conocer cuáles son las características de estas evaluaciones en una revista latinoamericana en crecimiento, se realizó un análisis de contenido a un conjunto de dictámenes entregados a la *Revista Colombiana de Psicología*. Específicamente, se analizaron las temáticas en las que los evaluadores se centraron más, en términos generales, dentro de tres categorías: (a) *aspectos formales*, (b) *aspectos editoriales* y (c) *aspectos de mérito científico*. Igualmente, se evaluó la medida en que las valoraciones hechas por el evaluador eran claras y constructivas.

Método

Muestra

Se analizaron 96 dictámenes evaluativos de 38 artículos diferentes, que fueron entregados a la *Revista Colombiana de Psicología* durante los primeros semestres de los años 2010 y 2011. Todos los artículos pasaron por un primer filtro de evaluación por parte del equipo editorial de la revista, antes de ser enviados a evaluadores. De estos dictámenes, el 28% corresponden a evaluaciones hechas a artículos teóricos (e. g. reflexiones sobre un tema y revisiones de literatura) y 72% son sobre artículos empíricos. Los dictámenes fueron redactados por evaluadores provenientes, principalmente, de países como Colombia (36.5%), España (24%), México (14.6%) y Brasil (9.4%).

Procedimiento

La recolección y análisis de los datos se realizó en tres etapas diferentes: (a) se construyeron las categorías de análisis, (b) a partir de estas categorías se determinó el porcentaje de texto dedicado por cada evaluador a cada una de estas categorías y (c) se realizó una evaluación del nivel en el cual las apreciaciones hechas por los evaluadores eran precisas y constructivas. A continuación se detallan cada una de estas etapas.

Construcción de categorías de análisis.

Se utilizó el software Atlas ti para analizar cada uno de los dictámenes y, a partir de una estrategia inductiva, se codificaron las apreciaciones hechas por los evaluadores. De esta forma, se construyeron diferentes categorías (en total 73) sobre los aspectos de los artículos de los cuales hablaban los evaluadores en sus dictámenes, por ejemplo, fallas metodológicas, errores de redacción y claridad en el problema de investigación, entre otros.

Al mismo tiempo, se definieron, a partir de la literatura revisada y la experiencia de los autores en el área editorial, tres grandes categorías

que abarcan las temáticas o aspectos que un jurado puede evaluar al revisar un manuscrito. Estas categorías son: (a) aspectos formales, que hacen énfasis, especialmente, en los contenidos referidos a la presentación del manuscrito; (b) aspectos editoriales, que involucran elementos relacionados con el contenido y la estructuración del texto, y se centran en la articulación y coherencia de las partes del manuscrito, y (c) mérito científico, en el cual se evalúa la relevancia del artículo para la disciplina y el fundamento científico del mismo. Una vez se establecieron estas categorías, se ramificaron en subcategorías, para lo cual se tuvo en cuenta las temáticas comunes en las 73 categorías que surgieron de la categorización realizada inicialmente.

Análisis de dedicación a categorías. Para evaluar la dedicación que los evaluadores dieron a cada una de las categorías, se contó el número de palabras que tenían las apreciaciones hechas en cada una de ellas. Posteriormente, se calculó el porcentaje de palabras que tenían las categorías con respecto al total de palabras empleadas en el dictamen y se compararon entre ellas. Para realizar esta comparación, se calculó la media de los porcentajes (siendo esta una variable continua) y se utilizaron estadísticos no paramétricos de comparación de medias para evaluar las diferencias. Luego se repitió este procedimiento con las subcategorías comparando, en esta ocasión, la cantidad de palabras de la subcategoría con respecto al total de la categoría a la que pertenecían.

Debido a que, como se verá más adelante, la categoría mérito científico incluye comentarios que tratan sobre aspectos metodológicos, para realizar el análisis de porcentajes en esta categoría, se excluyeron los dictámenes de artículos teóricos, de los cuales no se espera que tengan ese tipo de elementos.

Análisis de calidad de las apreciaciones. Finalmente, se analizaron las apreciaciones hechas por los evaluadores en las categorías

generales y se calificaron en dos aspectos específicos: claridad de las apreciaciones y grado en el que dichas apreciaciones fueron constructivas. Para evaluar lo primero, se tuvieron en cuenta las siguientes puntuaciones: se puntuaba con 1 cuando se infería que había una evaluación, pero no se conocía de manera precisa cuál; con 2, cuando se encontraba una afirmación evaluativa que no permitía conocer de forma precisa cómo se encontraba el artículo en ese aspecto (falla o fortaleza); y con 3, cuando había afirmaciones evaluativas que permitían conocer de forma precisa cómo se desarrollaba esa parte del artículo y que dejaban claro lo que llevaba al evaluador a realizar dicho juicio.

Para calificar el grado en que una apreciación era constructiva, se tuvieron en cuenta los siguientes criterios: se asignó 1 cuando en general los comentarios de una categoría no permitían conocer qué cambio específico debía realizar el autor en el artículo para mejorar el aspecto señalado y 2 cuando los comentarios explicitaban de manera concreta qué cambio debía realizar el autor para mejorar el artículo. Finalmente, a los comentarios en los que se hacía una evaluación positiva –lo que implicaba que no era necesario modificar el texto– se les asignó con el código N/A y fueron excluidos de los análisis porcentuales que se describen más adelante (para una mayor descripción de los criterios de calificación ver el Apéndice A).

Esta evaluación fue realizada por dos de los autores. El 33% de los dictámenes fue evaluado por ambos de manera independiente, obteniendo un índice de concordancia Kappa moderado de .54, para las calificaciones hechas sobre la claridad de las apreciaciones, y de .62 en la calificación del nivel en que los comentarios eran constructivos.

Finalmente, tras haber calificado cada uno de los dictámenes, se calcularon, por categoría, los porcentajes en los que cada uno de ellos obtenía alguna de las puntuaciones anteriormente descritas. Para calcular este porcentaje, se excluyeron los dictámenes que no tenían apreciaciones

en la categoría evaluada. Por ejemplo, si un dictamen no contenía apreciaciones sobre aspectos editoriales pero sí sobre el mérito científico, este se excluía del total de dictámenes que se usaban para calcular los porcentajes de la primera categoría, pero no de la segunda. Igualmente, para calcular el porcentaje de las apreciaciones que eran constructivas y de las que no, se excluyeron los casos en los que las valoraciones hechas no involucraban la necesidad de un cambio, es decir, que fueron calificadas con N/A.

Resultados

Definición de Categorías

Con base en lo que los evaluadores reportaron, se construyeron diferentes subcategorías que se organizaron dentro de las tres categorías generales descritas en la metodología: aspectos formales, aspectos editoriales y mérito científico. A continuación se describen cada una de las subcategorías y se dan ejemplos del tipo de citas textuales que se asociaron a cada una de ellas, con el objetivo de mostrar, de forma general, el tipo de contenido que los evaluadores tuvieron en cuenta en sus dictámenes.

Los aspectos formales se dividieron en cuatro subcategorías: (a) *uso del lenguaje*, (b) *seguimiento a las normas de presentación*, (c) *redacción y estilo*, y (d) *seguimiento de la política editorial*.

Uso del lenguaje. Esta categoría incluye los comentarios que se refieren al uso que los autores hacen de la gramática o al uso adecuado del lenguaje escrito, por ejemplo, la formación adecuada de frases, uso de signos de puntuación, concordancia de tiempos verbales, entre otros. A continuación, se presentan algunos ejemplos de comentarios de esta subcategoría:

“En la frase ‘esta situación se recrudece cuando se trata de cuidadores de pacientes con demencia (Izal & Montorio, 1999) especialmente tipo

alzheimer’, el nombre propio Alzheimer se escribe iniciando con la primera letra en mayúscula.”

“Es necesario revisar la ortografía, redacción y uso de expresiones poco técnicas.”

Seguimiento a las normas de presentación.

Incluye todos los comentarios que se refieren a las normas básicas de presentación establecidas por la revista, incluyendo las normas de publicación de la APA. Así, se encuentran comentarios relacionados con el tamaño de la fuente, interlineado, utilización de títulos y subtítulos, forma de las referencias y citaciones, extensión máxima de título, resumen y palabras clave, entre otros aspectos. Por ejemplo:

“Algunas referencias citadas no aparecen en el texto, otras no mencionan la fecha, algunas les falta organización, no ajustándose a nivel general a las normas APA.”

Redacción y estilo. Esta subcategoría, a diferencia de la categoría uso de lenguaje, va más allá de la correcta escritura y mira la medida en que las ideas se exponen de una forma clara y comprensible para el lector. Esto incluye comentarios sobre la repetición de ideas, la inclusión de ideas innecesarias, la extensión necesaria de un texto para exponer una idea y la utilización de un estilo de redacción científico, entre otros.

“Se sugiere modificar el penúltimo párrafo ya que es difícil seguir la descripción de la forma de responder en las cajas de respuesta.”

“Se utilizan algunos términos despectivos en apartes del trabajo, aunque la fuente original los utilice es importante ser críticos como autores al momento de la redacción y el análisis.”

Seguimiento de la política editorial. Tiene en cuenta todos los comentarios que hablan de la forma en que el manuscrito se adecúa o no al tipo de artículos que publica la revista, ya sea por su forma o su contenido. Por ejemplo, artículos de desarrollo teórico, revisiones de literatura y

reportes de investigación, entre otros. Un ejemplo de comentario para esta subcategoría es:

“El artículo tiene, más bien, el estilo de un ARTÍCULO DE DIFUSIÓN que podrían leer padres de familia, maestros, estudiantes, pero no corresponde con una revista científica, como la *Revista Colombiana de Psicología*.”

Por otro lado, la categoría de aspectos editoriales, se subdivide en tres subcategorías nuevas, que son: (a) *microestructura*, (b) *macroestructura* y (c) *descripción y organización de la macroestructura*. A continuación se describen cada una de estas subcategorías, dando, nuevamente, ejemplos de los tipos de comentarios que se encuentran en cada una de ellas.

Microestructura. Los comentarios hacen referencia a la manera en que fueron elaborados los elementos de la microestructura del texto, como el título, el resumen y las palabras clave, y hablan, principalmente, de la coherencia de estos elementos con el artículo en su conjunto. A diferencia de la subcategoría seguimiento de las normas de presentación, esta se centra más en el contenido de estas secciones, que en la sola presentación o en su aspecto formal. Un ejemplo de los comentarios que se encuentran en esta categoría es:

“El resumen ... no presenta de manera clara los resultados obtenidos o su relación con el tema en que se enmarca el estudio.”

Macroestructura. Los comentarios en esta categoría hablan de si los diferentes componentes generales que se esperan en un artículo científico están o no presentes en el manuscrito (e. g. la introducción, la revisión de la literatura, la pregunta de investigación, el método, los resultados, etc.) y tienen en cuenta si estos componentes son pertinentes para el tipo de artículo evaluado. Dos ejemplos de este tipo de comentarios son:

“En su INTRODUCCIÓN: no refleja a modo inicial los tópicos y objetivos de lo que será

abordado en su artículo, con vistas a crear una verdadera introducción.”

“Finalmente, sentí falta de una conclusión del autor/ de la autora encerrando la revisión.”

Descripción y organización de la macroestructura. Esta subcategoría habla de la medida en que los diferentes componentes del manuscrito (e. g. introducción, metodología, resultados, etc.), fueron realizados correctamente. En esta categoría el evaluador va más allá de si el elemento está o no presente en el manuscrito. Adicionalmente, se observa si estos componentes se encuentran ubicados en la sección en la que deberían estar y si tienen o no problemas de organización. Esta categoría incluye, también, el uso adecuado de ilustraciones (gráficos, figuras, tablas), donde se evalúa si se utiliza la herramienta visual que mejor se adecúa a la información que se quiere transmitir y si no se repite en el texto lo que se encuentra en las figuras o tablas, y viceversa. Un ejemplo de los comentarios que se encuentran en esta subcategoría es:

“Dado que el instrumento de violencia de pareja no ha sido publicado, debería describirse a profundidad, definiendo sus escalas, la forma en la que se validó y adjuntarlo al artículo.”

Finalmente, la categoría de mérito científico se dividió en tres subcategorías: (a) *aspectos conceptuales y constructos*, (b) *aspectos epistemológicos*, y (c) *aspectos metodológicos*.

Aspectos conceptuales y constructos. Esta subcategoría hace referencia a la claridad y precisión con la que se utilizan y definen los conceptos involucrados en el estudio. Son conceptos que se relacionan con la pregunta de investigación de una forma general y sirven de fundamento de la misma. Dos ejemplos de los comentarios que se encuentran en esta subcategoría son:

“En dicha introducción no se ofrece absolutamente ninguna información sobre la cuestión

de la violencia en la pareja ni en cuanto a qué se entiende como tal (por parte de los organismos internacionales que se ocupan de su análisis y abordaje o en el marco concreto de la psicología).”

“Falta precisión sobre las variables de estudio, se sugiere una definición más detallada de los constructos estrés y salud percibida.”

Aspectos epistemológicos. Esta subcategoría incluye los elementos que permiten que el estudio signifique un aporte para el conocimiento científico; tiene en cuenta la relevancia de la pregunta de investigación, los fundamentos del estudio y la forma en que el artículo, en su totalidad, es coherente con la pregunta de investigación. Esta categoría se subdividió en tres categorías más:

1. *Fundamentos:* esta categoría hace referencia al sustento empírico y teórico de la propuesta del artículo, incluye seleccionar referencias adecuadas y suficientes tanto para la pregunta de investigación como para la metodología utilizada para el desarrollo del proyecto. Por ejemplo:

“La introducción adolece de dos falencias importantes: (i) no distingue y revisa como análogas, o en gran medida, equivalentes, la exposición a la violencia en los medios de comunicación (léase TV), y la experiencia de jugar videojuegos con tramas y/o personajes violentos; (ii) La escasa y superficial consideración otorgada a las conclusiones de los estudios de Sherry (2001), Bensley y van Eenwyk (2001) y Sherry (2007).”

2. *Pregunta de investigación y objetivos:* esta categoría incluye comentarios relacionados con la relevancia de la pregunta, lo que implica que efectivamente constituya un aporte al campo de conocimiento en el que se enmarca. Por ejemplo:

“El tema en mención es interesante y de aporte al campo de la psicología de la salud, más sin embargo no es un tema de novedad ya que se han

escrito muchos artículos y capítulos de libro en los últimos 30 años sobre este argumento.”

3. *Resultados, discusión (o su equivalente en artículos teóricos) y conclusiones:* los comentarios hacen referencia a la medida en que estas secciones tienen una correcta estructura argumentativa y son coherentes con la pregunta de investigación, así como con los hallazgos a los que se llega. Los evaluadores indagan si la investigación efectivamente responde a la pregunta planteada y si con ella se logró o no hacer un aporte al conocimiento. Por ejemplo:

“¿Cuál es el fundamento empírico de la siguiente conclusión?: ‘... al tener un perro las personas realizan actividades que les ayudan a combatir el estrés y que contribuyen a su bienestar general ...’. Nuevamente, es necesario que los autores basen sus conclusiones en función de los resultados obtenidos mediante análisis estadísticos.”

Aspectos metodológicos. Esta subcategoría hace referencia a las apreciaciones hechas con respecto a la estrategia metodológica utilizada, donde se evalúa tanto la calidad del procedimiento utilizado como la adecuación de dicha estrategia para responder la pregunta de investigación. Igualmente, esta subcategoría se dividió en otras dos que se describen a continuación:

1. *Adecuación de la estrategia metodológica y el análisis de datos:* en esta subcategoría se incluyen comentarios sobre lo apropiado o no del método de investigación y del correspondiente análisis de los datos. Por ejemplo: “Se trata de la adaptación de una prueba norteamericana al español, y la adaptación-traducción de un test de unas culturas a otras requiere toda una tecnología y protocolos que los autores no han seguido.”
2. *Ejecución de la estrategia metodológica y el análisis de datos:* esta subcategoría involucra comentarios acerca de errores o fortalezas de

la ejecución misma de la estrategia de investigación, como el control de variables y la recolección de datos, entre otros. Por ejemplo: “No parece razonable utilizar el Modelo de Rasch de entrada y directamente, sin haber comprobado que los datos cumplen las condiciones exigibles para utilizar ese modelo logístico, por ejemplo la unidimensionalidad de los ítems.”

Distribución del Texto de Acuerdo con las Categorías Evaluadas

Comparación entre las categorías. La categoría en la que más se centraron los evaluadores fue la de mérito científico, con la media más alta ($\bar{x}=67.8$, $D.E.=24.4$), seguida por la categoría de aspectos editoriales ($\bar{x}=26.2$, $D.E.=19.5$) y, finalmente, la categoría de aspectos formales ($\bar{x}=16.5$, $D.E.=20.4$). Se realizaron pruebas de significancia, inicialmente con la prueba de Friedman y, posteriormente, la de Wilcoxon, para comparar cada categoría con la otra, y se encontró que todas las diferencias fueron significativas ($p < .01$).

Comparación entre las subcategorías. Al comparar entre las subcategorías se encontró en la categoría de aspectos formales, que la subcategoría de seguimiento de normas de presentación fue la que tuvo mayor porcentaje de palabras dedicadas, seguida por la de uso del lenguaje, redacción de estilo y, en último lugar, los comentarios sobre la política editorial (ver Tabla 1).

Tabla 1
Medias de porcentajes de palabras en las subcategorías de aspectos formales

Subcategoría	Media	D.E.
Uso de lenguaje	21.2	46.1
Seguimiento a normas de presentación	28.9	13.6
Redacción y estilo	14.5	36.1
Política editorial	1.69	8.8

Nota: los mínimos en todos los casos corresponden a 0.

Posteriormente, se realizó la comparación de los porcentajes de las cuatro subcategorías en las que se encontraron diferencias significativas ($\chi^2=52.6$, $p < .01$). Los análisis post-hoc mostraron que la única comparación en la que no hubo diferencias significativas fue entre las subcategorías de redacción y estilo y la del uso del lenguaje ($Z=1.6$, n.s.).

En cuanto a las comparaciones entre las subcategorías correspondientes a los aspectos editoriales, se encuentra que la descripción y organización de la macroestructura tienen la mayor media ($\bar{x}=86.5$, $D.E.=121.2$), seguida por los aspectos de microestructura ($\bar{x}=21.3$, $D.E.=53$) y, finalmente, se encuentran los aspectos de la macroestructura ($\bar{x}=16.4$, $D.E.=24.6$). Cuando se compararon las diferencias entre las tres medias, se hallaron diferencias significativas ($\chi^2=48.7$, $p < .05$). Las pruebas post-hoc evidenciaron que no existen diferencias significativas entre la subcategoría de macroestructura y microestructura ($Z=0.6$, n.s.), mientras que entre la descripción y organización de la macroestructura y las otras dos subcategorías, sí se encontraron diferencias significativas ($p < .01$).

Por último, se compararon las categorías correspondientes a las subcategorías de mérito científico, en las que se encontró que la categoría de aspectos epistemológicos tuvo la mayor media ($\bar{x}=73$, $D.E.=30$), seguida por la de aspectos metodológicos ($\bar{x}=17$, $D.E.=26$) y, por último, por la de aspectos conceptuales y constructos ($\bar{x}=6$, $D.E.=14$). Estas diferencias fueron significativas ($\chi^2=99.7$, $p < .01$) cuando se compararon todas las categorías entre sí, así como cuando se comparó cada una con las demás ($p < .01$).

Calidad de la Evaluación

Al comparar los porcentajes de comentarios que eran claros, por cada una de las categorías, se encontró que, en lo que respecta a los aspectos formales, la mayor parte de los evaluadores realizaron comentarios puntuados con

tres (63%), seguidos por comentarios puntuados con dos (32.4%) y, finalmente, por comentarios puntuados con uno (4.2%) ($n=71$). Estas diferencias resultaron ser significativas, según la prueba de Friedman ($\chi^2_2=34.5$, $p < .05$). En cuanto al carácter constructivo de los comentarios, se encontró que el 41.2% no permitían saber qué cambios específicos eran necesarios y el 58.8% daban alternativas precisas ($n=68$). La comparación estadística no arrojó diferencias significativas ($\chi^2_1=2.1$, n.s.).

Por otro lado, al evaluar la claridad de las apreciaciones de los aspectos editoriales se encontró que predominaron los comentarios puntuados con 3 (71.3%), seguidos de los comentarios puntuados con 2 (25.3%) y, finalmente, se encontraron los comentarios puntuados con 1 (3.4%) ($n=87$). Estas diferencias fueron significativas ($\chi^2_2=62.5$, $p < .05$). En cuanto a la calidad de las sugerencias, se encontró que 53.8% de los comentarios ofrecían alternativas precisas, mientras que un 46.2% no ofrecía alternativas para solucionar la dificultad ($n=78$). La comparación estadística muestra que estos porcentajes no son significativamente diferentes ($\chi^2_1=2.1$, n.s.).

Finalmente, en lo que respecta a la claridad de las apreciaciones sobre el mérito científico del manuscrito, se encontró que el mayor porcentaje de comentarios corresponde los puntuados con 3 (76%), seguidos por aquellos puntuados con 2 (19.8%) y, finalmente, se encontraron los que tenían un puntaje de 1 (3.1%) ($n=95$). La prueba estadística arroja diferencias significativas entre las tres categorías ($\chi^2_2=84.9$, $p < .05$). En cuanto al nivel en que la apreciación es constructiva, se encontró que en el 56.5% de los comentarios la alternativa que se sugiere es clara y en el 43.5% de los casos es vaga o no se realiza sugerencia alguna ($n=85$). No se encontraron diferencias significativas entre estos porcentajes ($\chi^2_1=1.4$, n.s.).

Discusión

Los evaluadores, cuyos conceptos fueron revisados en este estudio, prestaron más atención a los aspectos de mérito científico (que incluyen elementos epistemológicos, metodológicos y conceptuales) y menos a los aspectos formales de los manuscritos; es decir, se centraron en aquellos aspectos que especialmente ellos, como expertos en un área temática, podían evaluar y menos en los aspectos que otros actores involucrados en el proceso editorial podrían evaluar, por ejemplo, un corrector de estilo. Además, los comentarios o juicios hechos por los evaluadores fueron, en su mayoría, comentarios que permitían conocer de manera precisa la forma en que se encontraba el artículo en el aspecto que se evaluaba y que los llevó a realizar dicho juicio. Estos dos elementos, juntos, son de gran valor para el editor de una revista científica, pues le permiten tener más herramientas para juzgar (de manera fundamentada) la relevancia de un artículo para la comunidad académica y para el avance de la ciencia.

De igual manera, es de gran relevancia para el autor que los dictámenes que recibe describan, de manera concreta, cuáles fueron los elementos del manuscrito que llevaron al revisor a evaluarlo de determinada forma; el no hacerlo podría generar en el autor la sensación de que su trabajo no fue comprendido, además de producir en él emociones de rabia y frustración (Fletcher & Fletcher, 2003). Lo anterior, puede generar una reacción negativa del autor frente a la revista, que lo lleve, por ejemplo, a cuestionar el proceso de evaluación que esta realiza.

Por otro lado, se encontró que, dentro de las categorías generales, no hubo diferencias entre el número de comentarios constructivos y los no constructivos, es decir, que cerca de la mitad de las apreciaciones incluían sugerencias precisas acerca de qué cambios hacer en el artículo para mejorarlo y la otra mitad no incluían sugerencias de los posibles cambios, o las sugerencias

eran tan ambiguas que no se podía determinar cuál era el cambio específico que debía hacerse. Es importante resaltar este hecho pues, como se discutió en la introducción, el éxito de un autor al realizar los cambios que su trabajo requiere para ser publicado dependerá, en gran parte, de la claridad que tenga sobre cuáles son dichos cambios. Adicionalmente, si los evaluadores señalan los cambios específicos que se deben hacer en el manuscrito, será más sencilla la labor del editor a la hora de revisar un artículo corregido y determinar si se realizaron las modificaciones necesarias para que pueda ser publicado.

La diferencia que se observa entre la claridad de la evaluación y el carácter constructivo de los comentarios lleva a plantear la posibilidad de que la percepción del rol del evaluador esté dividida en dos: los que entienden que su papel es filtrar los artículos para que se publiquen únicamente aquellos que cumplan con los criterios necesarios para ello, o aquellos que, además de lo anterior, entienden que parte de su trabajo es dar una guía o instruir a los autores sobre las maneras particulares en las que su artículo puede ser mejorado. Esta idea es coherente con los hallazgos de Marušić et al. (2002), quienes igualmente identificaron estos dos tipos de evaluadores. En su estudio, ellos analizaron los dictámenes recibidos por dos revistas con grandes diferencias en la cantidad de artículos recibidos, en un periodo de año y medio. Encontraron que en las revistas con un menor flujo de artículos (269 artículos) los evaluadores y editores tenían una mayor tendencia a rescatar los elementos relevantes del artículo (e. g. resultados potencialmente relevantes) e invertir un esfuerzo para que los autores mejorarán dicho artículo. Por su parte, en las revistas con un mayor flujo de artículos (8,852 artículos) los evaluadores tenían una mayor tendencia a rechazar el artículo, en lugar de sugerir su publicación con cambios. Se esperaría que en la primera revista los jurados hicieran evaluaciones más constructivas, que le permitieran al autor mejorar su manuscrito,

mientras que esto no sería de relevancia para el caso de la segunda revista.

Como se ha hecho evidente a lo largo de este artículo, la labor del evaluador no va a afectar únicamente al editor, sino igualmente al autor que presenta el manuscrito para su publicación. Además, son varias las ocasiones en las que un autor realiza un estudio empírico, que responde a una pregunta de investigación relevante en el área, pero no es exitoso al redactar el informe de investigación o artículo científico. Es especialmente importante en estos casos que tanto el editor como el evaluador centren sus esfuerzos en rescatar dicho trabajo y le den al autor las herramientas para hacer un buen artículo. Jefferson, Wager y Davidoff (2002), basados en una amplia revisión de literatura, dividen en dos los objetivos de los evaluadores: el primero es seleccionar los artículos para publicación y descartar los manuscritos con contenido irrelevante, trivial, erróneo o potencialmente dañino; el segundo es mejorar la precisión, la claridad, la transparencia, la exactitud, la validez y la utilidad de los manuscritos seleccionados. Sin embargo, Kumar (2009), a partir de una revisión sobre las problemáticas en la evaluación por pares o jurados, concluye que no hay un consenso respecto de si los evaluadores están o no cumpliendo con ambos objetivos.

Finalmente, en el presente trabajo se encontró que, al comparar las tres categorías generales (aspectos formales, editoriales y de mérito científico), no hubo diferencias entre la claridad de los comentarios ni en el grado en que los comentarios eran constructivos o no. Esto significa que los comentarios eran del mismo tipo, independientemente del aspecto que se estuviera evaluando.

Estos tres aspectos, la evaluación que hacen los jurados del contenido del manuscrito, el grado en que los comentarios hechos son claros y concretos, y el grado en que son constructivos, son elementos que otros autores también han resaltado como relevantes dentro de la labor que hacen los evaluadores. Schroter, Tite, Hutchings

y Black (2006), por ejemplo, utilizaron un instrumento para evaluar la calidad de la revisión que preguntaba si el evaluador se enfocaba en aspectos como la pregunta de investigación, la originalidad del trabajo o los aspectos metodológicos, entre otros. Pero, igualmente, el instrumento preguntaba si el evaluador realizaba comentarios constructivos y si daba evidencia apropiada, utilizando ejemplos del manuscrito para sustentar sus comentarios.

De estos tres aspectos, el que requiere mayor atención (debido a que fue el elemento más débil en las evaluaciones) es la realización de comentarios constructivos que digan de manera precisa qué cambio debe realizar el autor y dónde. En esta línea, podría ser importante que los editores expongan de manera clara, en las comunicaciones que tengan con los evaluadores, la importancia de cumplir con este criterio en sus evaluaciones. Sin embargo, hay que hacer la salvedad de que hacer comunicaciones más detalladas a los evaluadores respecto de lo que se espera de ellos, no garantiza que lo hagan, por lo que los efectos de esta estrategia sobre la labor del evaluador son aún un problema a investigar (Fletcher & Fletcher, 2003).

Atención Prestada por los Evaluadores a cada una de las Subcategorías

A partir del análisis de contenido, realizado de manera inductiva, y de las tres categorías generales elaboradas deductivamente se construyeron diez subcategorías principales, cuatro de la categoría aspectos formales (que incluye uso de lenguaje, seguimiento a las normas de presentación, redacción y estilo, y seguimiento de la política editorial), tres subcategorías de aspectos editoriales (que incluye microestructura, macroestructura y descripción y organización de la macroestructura), y tres de la categoría mérito científico (que incluye aspectos conceptuales y constructos, aspectos epistemológicos y aspectos metodológicos).

Al revisar las subcategorías del aspecto formal, se encontró que los evaluadores se centraron más en el seguimiento de normas, seguido por los aspectos de uso del lenguaje y redacción y estilo (no hubo diferencias significativas entre estos dos) y, finalmente, en los comentarios referidos a la política editorial. Al analizar este resultado, hay que recordar que todos los artículos fueron evaluados, antes de enviarse a revisión por pares, por el equipo editorial de la revista, que determinaba, entre otras cosas, si el contenido y formato del manuscrito se adecuaba al estilo y a las políticas de la revista. En este sentido, se esperaría que ninguno o muy pocos de los evaluadores consideraran que alguno de los artículos no cumplía con la política editorial de la revista.

En la categoría de aspectos editoriales, se encontró que la subcategoría a la que los evaluadores le dieron mayor relevancia fue a la de descripción y organización de los elementos de la macroestructura, y que se centraron menos en los aspectos de la microestructura y la macroestructura. Estas dos últimas subcategorías incluyen comentarios sobre la adecuación de elementos básicos como el título, el resumen y las palabras clave, o la presencia o ausencia de una de las secciones principales del artículo (e. g. introducción, metodología, resultados, etc.); mientras que la categoría de descripción y organización de los elementos de la macroestructura va más allá y mira qué tan bien descritos están estos datos y si su organización es la adecuada, es decir que esta categoría es la más relacionada con el contenido del artículo. En este sentido, que los evaluadores se centren en esta última es acorde con las necesidades de los editores y con lo que se espera que los evaluadores realicen y evalúen. Las otras dos categorías, nuevamente, podrían ser revisadas por un profesional que no necesariamente tenga la formación y el nivel académico que usualmente tienen los evaluadores de una revista científica (i. e. una formación de posgrado).

Por otro lado, hay que tener en cuenta que los criterios de evaluación enviados a los jurados de la muestra se centraban, especialmente, en los aspectos que se incluyen dentro de la subcategoría descripción y organización de los elementos de la macroestructura. Estos criterios incluían cada una de las partes que se esperan de un artículo científico (que variaba dependiendo de si el artículo era empírico o teórico) y una descripción de qué tipo de información se esperaba en cada una de estas secciones. Por ejemplo, el criterio de evaluación mencionaba que la introducción debía presentar los antecedentes en el área, así como plantear y justificar de manera clara la pregunta de investigación; también mencionaba que en la metodología se deben describir los instrumentos utilizados y sus propiedades psicométricas. En este sentido, se podía esperar que varios de los comentarios hechos por los evaluadores se centraran en los aspectos que se señalan en los criterios de evaluación.

Finalmente, al evaluar la dedicación a las subcategorías de la categoría de mérito científico, se encontró que la subcategoría en la que más centraron su atención fue el aspecto epistemológico, es decir, todos aquellos aspectos relacionados con el sustento empírico y teórico de la pregunta, la relevancia del problema de investigación y la coherencia de los resultados, discusión y conclusiones respecto a la pregunta. En términos generales, esta categoría tiene que ver con el aporte del artículo al avance en el conocimiento. Estos resultados coinciden con los de Bornman et al. (2008), quienes en su metaanálisis encontraron que el criterio de evaluación más empleado por los evaluadores fue la relevancia de la contribución.

Si se tienen en cuenta las diez subcategorías, la subcategoría del aspecto epistemológico sería la más recurrente de todas. Esto resulta bastante positivo, dado que los editores valoran en gran medida si un artículo representa un avance para el conocimiento. Esto se ve reflejado en algunos

estudios, por ejemplo, Wolff (1970) entrevistó a 66 editores que consideraron que son tres los elementos más importantes de un artículo: que signifique un avance en el conocimiento, que tenga una metodología adecuada y que los resultados se presenten de manera objetiva. Por su parte, Bornman et al. (2008) encontraron que los aspectos más valorados por los editores y evaluadores en la revisión de un manuscrito son la discusión de los resultados (mirar si las conclusiones del manuscrito son objetivas, correctas y basadas en los resultados), que exista un avance teórico, un diseño y una concepción (marco conceptual adecuado y un correcto diseño de la investigación), y una adecuación de la metodología y el análisis de datos.

Como muestran estos autores, otro aspecto de gran relevancia para los editores, aparte del avance en el conocimiento, es el correcto uso de las estrategias metodológicas; sin embargo, en la muestra estudiada, los aspectos metodológicos fueron mencionados en un porcentaje mucho menor que los epistemológicos. Este es un elemento que requiere de especial atención, ya que analizar superficialmente los aspectos metodológicos y de análisis de datos puede llevar a publicar artículos con errores graves en su metodología. Por ejemplo, Campanario (2002) describe algunos estudios que observaron que cuando se revisaban una segunda vez artículos que fueron originalmente publicados, se encontraron errores estadísticos graves. Estos errores podrían invalidar por completo las conclusiones y análisis de dichos artículos que, sin embargo, se encuentran circulando por la comunidad académica. Un lector desprevenido podría omitir estos errores y continuar reproduciendo un conocimiento que no estaría fundamentado.

Ahora bien, un punto para reflexionar sobre este aspecto es la dificultad que puede significar que un evaluador sea experto tanto en la temática del artículo como en la estrategia metodológica empleada; eso teniendo en cuenta el aumento en la variedad de estrategias

metodológicas y estadísticas en el área de la salud (Altman & Schulz, 2003). Un editor podría utilizar gran parte de sus esfuerzos en encontrar a un revisor que maneje ambas temáticas, sin embargo, en varias ocasiones resulta difícil encontrar a un académico experto en una de las temáticas, por lo que buscar uno que sea experto en ambas puede multiplicar esa dificultad. En esta línea, se podrían enviar los artículos a un experto en la temática y a un experto en la metodología, y aclarar a cada uno los aspectos en los que se espera que se centren.

Finalmente, la subcategoría de la categoría de mérito científico a la que se hizo menos referencia fue la de aspectos conceptuales y constructos. Se refiere a la claridad y la precisión con la que se utilizan y definen los conceptos involucrados en el estudio.

Esta distribución (un fuerte énfasis en los aspectos epistemológicos y poco énfasis en los metodológicos y conceptuales) lleva a plantear la idea de que la comunidad académica se centra en la generación de nuevo conocimiento, pero presta menor atención a los fundamentos conceptuales y empíricos a partir de los cuales se obtienen dichos resultados. Un resultado sin un fundamento de este tipo no es útil, por lo que vale la pena llamar nuevamente la atención de los académicos, y especialmente de los evaluadores, sobre este aspecto.

Conclusiones

El papel del evaluador es fundamental en la publicación de artículos científicos de calidad; debe orientar su labor no solo a seleccionar los artículos que cumplan con este criterio, sino también a dar las herramientas al autor para que pueda mejorar su manuscrito, especialmente cuando hay un trabajo empírico importante de fondo. En esta línea, la labor del evaluador deberá ser útil para el editor y para el autor.

Claramente, los evaluadores de la muestra realizaron una labor que es acorde con las necesidades del editor; se centraron en los aspectos

que, especialmente ellos como expertos en un área, pueden evaluar y que los editores valoran en gran medida a la hora de tomar una decisión (por ejemplo, el avance en el conocimiento). Además, los comentarios de los jurados en su mayoría fueron precisos, es decir, mostraron la manera concreta en que se encontraba el artículo en el aspecto evaluado. Se observó que los evaluadores se centraron en mayor medida en los aspectos que hablan de la construcción de nuevo conocimiento y menos en los aspectos metodológicos del manuscrito. Este último aspecto es altamente valorado por los editores a la hora de decidir frente a la publicación de un manuscrito, ya que es fundamental para la validez de los resultados y las conclusiones de un estudio, por lo que se sugiere a los evaluadores prestarle mayor atención.

Por otro lado, casi la mitad de las valoraciones hechas por los evaluadores servían para las necesidades de los autores y la otra mitad no: solo la mitad de ellas hacían comentarios constructivos que mostraban de manera clara las modificaciones que debía hacer el autor para mejorar su texto. Como se dijo anteriormente, la función del evaluador no es solo seleccionar los artículos que deben ser publicados, sino también dar las herramientas para que el manuscrito mejore su calidad. Esto es especialmente importante en países donde el nivel de producción científica (artículos, revistas científicas, etc.) es reducido, pero está en crecimiento, como es el caso colombiano y latinoamericano. Es importante que los pares revisores tengan esto en cuenta a la hora de realizar sus evaluaciones, de esta forma su aporte doblaría el que ya le brindan a la comunidad académica.

Finalmente, es importante mencionar que parte del análisis que se presentó en este artículo se basa en un conteo de palabras en cada una de las subcategorías y en el porcentaje de palabras respecto al total de palabras en cada una de las categorías que se construyeron en el

proceso de análisis de contenido. Sin embargo, es posible que la naturaleza de los comentarios que se hacen sobre el mérito científico, por ejemplo, requieran siempre utilizar un mayor número de palabras que los comentarios sobre forma. En esta medida, las diferencias en cantidad pueden no corresponderse necesariamente con las diferencias en la relevancia dada por los evaluadores, sino con una diferencia dada por la naturaleza y el contenido de los comentarios. Por esto, sería relevante hacer un análisis empírico que confirme esta suposición (los comentarios, dependiendo de su naturaleza, requerirán siempre más o menos palabras), para responder a esa dificultad y profundizar en los resultados aquí expuestos.

Una alternativa para solucionar esta dificultad hubiera sido realizar el conteo del número de citas por categoría, lo cual hubiera medido de otra manera la relevancia dada por los revisores a cada una de las categorías. Sin embargo, en el presente estudio se optó por utilizar como unidad de medida el porcentaje de palabras, en lugar de la cantidad de citas en una misma subcategoría, debido a que esta última no resultó una medida adecuada para comparar los dictámenes. Esto se debe a que un evaluador podía realizar varios comentarios aislados para hacer referencia a una misma idea, o describir la misma idea en un solo comentario. Si se cuenta el número de citas en el primero de los casos se encontrarían más citas que en el segundo, y esto podría haber llevado a la idea errada de que el primer caso se centró más en ese aspecto particular que el segundo (al tener más afirmaciones sobre la misma idea), cuando la dedicación pudo o no haber sido la misma.

Referencias

- Altman, D. G. & Schulz, K. F. (2003). Statistical peer review. En F. Godlee & T. Jefferson (Eds.), *Peer Review in Health Sciences* (2nd ed., pp. 191-208). London: BMJ Publishing Group.
- Benos, D. J., Kirk, K. L. & Hall, J. (2003). How to review a paper. *Advances in physiology education*, 27 (2), 47-52.
- Bornmann, L., Nast, I. & Daniel, H.-D. (2008). Do editors and referees look for signs of scientific misconduct when reviewing manuscripts? A quantitative content analysis of studies that examined review criteria and reasons for accepting and rejecting manuscripts for publication. *Scientometrics*, 77, 415-432.
- Buela-Casal, G. & López, W. (2005). Evaluación de las revistas científicas iberoamericanas. Iniciativas y estado actual. *Revista Latinoamericana de Psicología*, 37, 211-217.
- Campanario, J. M. (2002). El sistema de revisión por expertos (peer review): muchos problemas y pocas soluciones. *Revista Española de Documentación Científica*, 25, 267-285.
- Fletcher, R. H. & Fletcher, S. W. (2003). The effectiveness of journal peer review. En F. Godlee & T. Jefferson (Eds.), *Peer Review in Health Sciences* (2nd ed., pp. 62-75). London: BMJ Publishing Group.
- Guerrero, J. & Jaraba, B. (s.f.). *La producción científica de la psicología colombiana: un análisis bibliométrico de las revistas académicas, 1949-2008*. Recuperado de http://www.ascofapsi.org.co/observatorio/documentos/2011/Estudio_Bibliometria.pdf
- Jefferson, T., Wager, E. & Davidoff, F. (2002). Measuring the quality of editorial peer review. *Journal of American Medical Association*, 287, 2786-2790.
- Kumar, M. (2009). A review of the review process: Manuscript peer-review in biomedical research. *Biology and Medicine*, 1, 1-16.
- Marušić, A., Lukić, I., Marušić, M., McNamee, D., Sharp, D., & Horton, R. (2002). Peer review in a small and a big medical journal: Case study of the Croatian medical journal and the lancet. *Croatian Medical Journal*, 43, 286-289.
- Rowland, F. (2002). The peer-review process. *Learned Publishing*, 15, 247-258.
- Schroter, S., Black, N., Evans, S., Godlee, F., Osorio, L. & Smith, R. (2008). What errors do peer reviewers detect, and does training improve their ability to detect them? *Journal of the Royal Society of Medicine*, 101, 507-514.

- Schroter, S., Tite, L., Hutchings, A. & Black, N. (2006). Differences in review quality and recommendations for publication between peer reviewers suggested by authors or by editors. *Journal of the American Medical Association*, 295, 314-317.
- Trzesniak, P. (2009). A estrutura editorial de um periódico científico. En A. A. Sabadini, M. I. Sampaio & S. H. Koller (Eds.), *Publicar em psicologia um enfoque para a revista científica* (pp. 88-101). São Paulo: Associação Brasileira de Editores Científicos de Psicologia.
- Vera-Villaruel, P., López-López, W., Lillo, S. & Silva, L. (2011). La producción científica en psicología latinoamericana: un análisis de la investigación por países. *Revista Latinoamericana de Psicología*, 43, 95-104.
- Wolff, W. M. (1970). A study of criteria for journal manuscripts. *American Psychologist*, 25, 636-639.
- Zych, I. & Buela-Casal, G. (2010). Internacionalidad de las revistas de psicología multidisciplinar editadas en Iberoamérica e incluidas en la Web of Science. *Universitas Psychologica*, 9, 27-34.

Apéndice A

Escalas de Puntuación para Calificar la Precisión de la Evaluación y de las Alternativas de Corrección

Escala de Puntuación para la Claridad de los Comentarios Evaluativos

Se puntúa 1 cuando se infiere que hay una evaluación, pero no se conoce de manera precisa cuál. Por lo general, esta calificación se asigna a comentarios que hacen referencia a posibles modificaciones en el texto, sin embargo, no es claro cuál es la dificultad que presenta esta parte del artículo y por qué requiere una modificación. Por ejemplo:

“Precisamente, los 30 extractos musicales presentados a cada participante, tenían una duración de 5 segundos. Esa es la duración que presentan gran parte de los sonidos incluidos en el IADS (otros duran 6 segundos). Por tanto, en investigaciones futuras quizá sería interesante utilizar los sonidos del IADS (con la posible manipulación de su tempo), ya que están baremados en las dos dimensiones emocionales empleadas en la presente investigación.”

Se puntúa 2 cuando se encuentra una afirmación evaluativa que no permite conocer de forma precisa cómo se encuentra el artículo en ese aspecto (falla o fortaleza). Esto ocurre especialmente cuando se hace una apreciación del artículo que resulta muy general, por lo que no se comprende por qué el evaluador la hace. Por ejemplo:

“Los resultados presentados presentan información redundante, lo cual crea ruido a la hora de comprender los mismos.”

Se asigna 3 cuando se encuentran afirmaciones evaluativas que permiten conocer de forma precisa cómo se desarrolla esta parte del artículo y que dejan claro lo que lleva al evaluador a realizar dichas observaciones. Por ejemplo:

“Los autores concluyen que las personas mayores muestran un deterioro en el reconocimiento tanto en codificación dividida como en recuperación dividida. Esta conclusión se ofrece porque no hay diferencias entre las dos condiciones en los ancianos. Sin embargo, sin una condición control en que tanto la codificación como la recuperación se realicen en condiciones normales (sin división de atención) es difícil saber si la ausencia de diferencia se debe a que no hay efectos de atención dividida en los ancianos o a que el efecto de atención dividida es igual para codificación y recuperación.”

Escala de Clasificación para Calificar el Grado en que una Apreciación es Constructiva

Se asignó 1 cuando el comentario no permitía conocer qué cambio específico debía realizar el autor en el artículo para mejorar el aspecto señalado. Por ejemplo:

“La discusión intenta responder a las preguntas de investigación, aunque en este ítem no se aprecia diferencias sustanciales con los resultados de autores como D. Kuhn sobre el mismo tema.”

“La discusión debería haber incorporado más autores para enriquecer el texto.”

Se asignó 2 cuando los comentarios explicitaban de manera concreta qué cambio debía realizar el autor para mejorar el artículo. Por ejemplo:

“La poca información dada en el artículo sobre la naturaleza del entrenamiento recibido por los estudiantes, impide al lector reflexionar más profundamente sobre posibles implicaciones educativas del estudio. Una discusión de la naturaleza específica de ese entrenamiento, a la luz de los resultados obtenidos sería, por tanto, bienvenida¹.”

1 Cita original en portugués: “A pouca informação dada no artigo sobre a natureza do treinamento recebido pelos estudantes, impede o leitor de refletir mais profundamente sobre possíveis implicações educacionais do estudo. Uma discussão da natureza específica desse treinamento, à luz dos resultados obtidos seria, portanto, bem-vinda”.

Finalmente, a los comentarios en los que se hace una evaluación positiva, lo que implica que no es necesario modificar el texto, se les asignó el código N/A y se excluyeron de los análisis porcentuales. Por ejemplo:

“El tema presentado sin duda constituye una importante reflexión que puede contribuir a que en Latinoamérica se avance en la comprensión y el uso de la TRI, contribuyendo de esta manera al avance en medición psicológica.”