



REICE. Revista Iberoamericana sobre Calidad,
Eficacia y Cambio en Educación

ISSN: 1696-4713

rinace@uam.es

Red Iberoamericana de Investigación Sobre Cambio
y Eficacia Escolar
España

Escala para Evaluar Artículos Científicos en Ciencias Sociales y Humanas- EACSH

López-López, Emerson; Tobón, Sergio; Juárez-Hernández, Luis Gibrán

Escala para Evaluar Artículos Científicos en Ciencias Sociales y Humanas- EACSH

REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, vol. 17, núm. 4, 2019

Red Iberoamericana de Investigación Sobre Cambio y Eficacia Escolar, España

Disponible en: <https://www.redalyc.org/articulo.oa?id=55166992007>

Escala para Evaluar Artículos Científicos en Ciencias Sociales y Humanas- EACSH

Scale to Evaluate Scientific Articles in Social and Human Sciences- SSAHS

*Emerson López-López **

CIFE, México

emerson.lopez@ulv.edu.mx

 <https://orcid.org/0000-0001-6149-7856>Redalyc: [https://www.redalyc.org/articulo.oa?](https://www.redalyc.org/articulo.oa?id=55166992007)

id=55166992007

Sergio Tobón

CIFE, México

stobon@cife.edu.mx

 <https://orcid.org/0000-0001-5543-9131>*Luis Gibrán Juárez-Hernández*

CIFE, México

luisgibran@cife.edu.mx

 <https://orcid.org/0000-0003-0658-6818>

Recepción: 19 Marzo 2019

Aprobación: 04 Julio 2019

RESUMEN:

En el mundo académico sorprende la baja cantidad de artículos científicos que son aceptados a investigadores nóveles, debido, en parte, a que faltan elementos importantes a nivel formal o de contenido. Además, son escasos los instrumentos debidamente validados que faciliten la escritura de los artículos científicos, de modo que el autor novel pueda autoevaluar su avance. Por ello, se creó la Escala para Evaluar Artículos Científicos en Ciencias Sociales y Humanas-EACSH, mediante un estudio exploratorio basado en la revisión de la literatura, el análisis de expertos en publicaciones científicas, la validación de contenido y la validez de constructo y confiabilidad, por medio de un muestreo censal con 129 investigadores de alto nivel de México durante el año 2018. Los ítems de la escala se validaron en términos de pertinencia y redacción (V de Aiken $> 0,80$), en donde más del 90% de estos estuvieron representados en el modelo factorial. El instrumento presentó una confiabilidad óptima (0,937) y quedó conformado por 19 ítems, demostrando que puede ser apropiado para guiar la escritura y evaluación de artículos científicos en el contexto Latinoamericano, cuyo contenido sean investigaciones de tipo cuantitativo con alcance exploratorio y descriptivo, aplicables a las ciencias sociales y humanas.

PALABRAS CLAVE: Análisis factorial, Base de datos, Evaluación, Publicaciones, Escritura.

ABSTRACT:

In the academic world, the low number of scientific articles that are accepted by new researchers is surprising, due, in part, to the lack of important elements at a formal or content level. In addition, there are few duly validated instruments that facilitate the writing of scientific articles, so that the novel author can self-assess their progress. Therefore, the Scale to Evaluate Scientific Articles in Social and Human Sciences-SSAHS was created, through an exploratory study based on literature review, analysis of experts in scientific publications, content validation and construct validity and reliability, through a census sampling with 129 high-level researchers from Mexico during 2018. The items of the scale were validated in terms of relevance and writing (V of Aiken $> 0,80$), where more than 90% of these were represented in the factorial model. The instrument presented an optimal reliability (0,937) and was composed of 19 items, demonstrating that it may be appropriate to guide the writing and evaluation of scientific articles in the Latin American context, whose content is quantitative research with exploratory and descriptive scope, applicable to the social and human sciences.

NOTAS DE AUTOR

* Contacto: emerson.lopez@ulv.edu.mx

KEYWORDS: Factorial analysis, Data base, Evaluation, Publications, Writing.

1. REVISIÓN DE LA LITERATURA

En investigación, lo que no está escrito no existe y menos aún si no está disponible en la web (Murillo, Martínez-Garrido y Belavi, 2017). Esta es una afirmación por demás preocupante y a la vez de interés para el mundo de los investigadores de todas las áreas del conocimiento. Más todavía, para el caso latinoamericano donde las publicaciones de corte educativo en revistas de alto impacto por autores de la región son escasas (Murillo y Martínez-Garrido, 2019), lo cual se traduce en la urgente necesidad de brindar más y mejores apoyos para la redacción de artículos científicos a los distintos sectores y actores del conocimiento.

Sin embargo, la fuerte competitividad que existe entre los artículos científicos que son sometidos a las revistas indexadas para su posible publicación, es innegable debido a que entre el 80% y 95% o más de los artículos científicos sometidos a las revistas de alto impacto en el mundo son rechazados; siendo solo un porcentaje mínimo el aceptado para publicación (LaPlaca, Lindgreen y Vanhamme, 2017). Esto significa que la competencia para escribir y publicar artículos de esta naturaleza es de alta exigencia, razón por la que se hace indispensable orientar con precisión a los investigadores noveles sobre la elaboración de artículos científicos.

Pero, ¿qué es un artículo científico? Entre la diversidad de definiciones, se opta en este escrito por definirlo como el documento oficial que la comunidad científica ha determinado para comunicar resultados de investigaciones entre sus miembros, cuyas características estructurales se dan en el siguiente orden: título, autores, resumen y palabras clave, *abstract* y *keywords*, introducción, metodología, resultados, discusión (denominados estos cuatro últimos como estructura IMRYD), agradecimientos y lista de referencias (APA, 2010; Pereira, 2014; Santesteban y Núñez, 2017).

Dichas características estructurales no siempre están presentes en todos los artículos dada la naturaleza de estos, pero sí los elementos centrales (IMRYD) que le dan coherencia y solidez a la investigación, aunque con leves variantes dependiendo de diversos factores, que pueden detallar o sintetizar dichos elementos por política de presentación del artículo en determinada revista o por otras razones ajenas que no demeritan el contenido del estudio que es lo sustancial de todo buen trabajo investigativo.

Para evaluar si pueden o no ser publicables estos artículos, primero, el autor puede autoevaluarlo mediante listas de cotejo o escalas de estimación construidas por las revistas a las cuales desea someter el artículo para su publicación; ya estando en manos de la revista, la siguiente evaluación la hacen los editores y los árbitros que colaboran con ella (Debnath y Venkatesh, 2015). Una vez publicado el artículo, la comunidad científica ha aceptado la valoración de la calidad mediante el factor de impacto como el Journal Citation Reports (JCR), el índice H de Hirsch, el índice G, entre otros, que miden en qué grado el artículo está siendo citado por autores de otros artículos (UPGC, 2018).

Sin embargo, las listas de cotejo y escalas de estimación que emplean las revistas científicas pocas veces tienen estudios de validez y confiabilidad. Son pocos los estudios en torno a la validez y confiabilidad de escalas para evaluar artículos científicos, y las que hay generalmente se centran en aspectos formales, sin profundizar en elementos que ayuden a determinar el impacto y la pertinencia de la investigación. En la literatura se han propuesto algunos instrumentos validados para evaluar un artículo científico (Greenberg, 2015; Merma, Peña y Peña-Alfaro, 2017; Rakedzon y Baram, 2017), pero tienen problemas tales como: 1) les falta abordar todos los elementos de un artículo paso a paso; 2) algunos indicadores no son claros; y 3) a veces sus descriptores son generales y poco concretos, lo cual dificulta su empleo en la evaluación y mejora de un artículo científico.

Por lo anterior, es necesario elaborar escalas integrales que apoyen la redacción de los artículos científicos en los autores noveles, editores y revisores, a tal grado que al someterse a revistas indexadas tengan mayor probabilidad de ser aceptados. Por ello, en la presente investigación, se buscó lograr las siguientes metas: 1)

elaborar una escala para evaluar los ejes esenciales de un artículo científico en las áreas de las ciencias sociales y humanas, tal y como lo proponen las normas APA Sexta Edición (APA, 2010), con énfasis en la metodología exploratoria y descriptiva, 2) determinar la validez de contenido de la escala a partir de un grupo de expertos en el área; 3) someter la escala a un grupo piloto para valorar su desempeño y pertinencia en las áreas de las ciencias sociales y humanidades, 4) establecer la validez de constructo del instrumento, y 5) determinar la confiabilidad de la misma.

2. MÉTODO

Enfoque

El enfoque metodológico en términos generales, es cuantitativo, de alcance descriptivo, no experimental, transversal y desarrollado mediante la lógica deductiva. Se efectuó un estudio instrumental, el cual atiende el diseño de instrumentos de medición para estudiar las propiedades psicométricas de los mismos, o adaptarlos a otros contextos geográficos y socioculturales (Montero y León, 2005, 2007), en torno a la validez de contenido, constructo y confiabilidad de una escala para evaluar y mejorar un artículo científico. Para este proceso, primero se hizo una amplia búsqueda de literatura referente a cómo se construyen artículos científicos de alcance exploratorio y descriptivo en las áreas de ciencias sociales y humanas; al mismo tiempo se buscaron instrumentos validados en bases de datos como Web of Science y Scopus. Recopiladas las fuentes pertinentes, se procedió a realizar una lectura analítica de las mismas, revisando a la vez las escalas encontradas y estableciendo sus aspectos comunes además de sus vacíos, cotejándolas al mismo tiempo con las normas APA Sexta Edición, sobre cómo debe elaborarse un artículo científico. El diseño metodológico tuvo tres fases: fase 1. Construcción del instrumento; fase 2. Estudio piloto; fase 3. Validez de contenido; y fase 4. Validez de constructo y confiabilidad.

Construcción del instrumento

Para construir el instrumento se siguió un procedimiento similar al realizado en otras investigaciones recientes sobre la validez de instrumentos, como las reportadas por González (2018), Tobón y colaboradores (2018) y Vincés (2018). Se determinó en primer lugar el constructo a evaluar: un artículo científico, el cual fue definido operacionalmente como un escrito relativamente sintético que posee un estilo de redacción y estructura previamente definida por la comunidad científica, cuyo contenido pasa antes de su publicación por el filtro de la evaluación arbitrada. El contenido principal gira en torno a las características de la variable estudiada, las cuales son transferidas a datos estadísticos y se comunican de manera clara y concisa a los académicos o científicos. Posteriormente se procedió a elaborar la escala con ocho ejes o dimensiones, las cuales fueron descritas siguiendo la estructura de un artículo. Una vez estructurada la escala, tres expertos con grado de doctor en educación y con experiencias en diseño y validación de instrumentos la revisaron, y con base en su retroalimentación fue mejorada, quedando como se observa en el cuadro 1. Cada uno de los 21 ítems se evalúa con una escala de cinco puntos: 1=nivel muy bajo; 2=nivel bajo; 3=nivel medio; 4=nivel medio alto; y 5=nivel muy alto.

CUADRO 1
Estructura de la escala

DIMENSIONES	DESCRIPCIÓN DE CONTENIDO E INDICADORES
I. Portada y Resumen	<p>Es la presentación del artículo y se compone del título, nombre de los autores, afiliación institucional, resumen y <i>abstract</i>, palabras clave y <i>keywords</i>.</p> <ol style="list-style-type: none"> 1. El título describe el tema esencial del artículo, es sencillo y claro, atractivo y breve, tiene menos de 16 palabras, está centrado y las palabras de más de cuatro letras empiezan con mayúscula. 2. Se indican los nombres de los autores después del título, los apellidos van unidos con guion (a menos que se trate de un único apellido), se agrega la afiliación institucional y el correo electrónico de cada autor. 3. El resumen está en español e inglés y se presenta con máximo 250 palabras (o el número de palabras permitido por la revista en la cual se espera publicar). Además, describe brevemente el problema, objetivo, metodología, principales resultados y conclusiones del estudio. 4. El número de palabras clave está entre 4 y 8 (o dentro del rango que permite la revista en la cual se espera publicar), fueron extraídas de un tesoro de la disciplina, están escritas con minúscula, separadas con coma y en orden alfabético.
II. Introducción	<p>Es la presentación de los antecedentes del tema, formula y explica el problema, describe la importancia de la investigación y presenta los propósitos del estudio.</p> <ol style="list-style-type: none"> 5. Se inicia con algunos párrafos atractivos que presentan la relevancia, justificación, y el contexto del tema, centrandolo al lector en el problema de estudio y motivando a la lectura del artículo. 6. Se hace una revisión crítica sobre los principales estudios realizados sobre el problema, considerando los propósitos del estudio, desde lo más general a lo particular. Además, se basa en citas parafraseadas en la mayoría de los casos, integrando información reciente con información histórica. 7. Los objetivos son relevantes, se relacionan con el problema de investigación expuesto y están redactados con claridad (poseen un verbo en infinitivo, objeto, sujeto y contexto del estudio).
III. Metodología	<p>Consiste en el proceso a través del cual se hizo el estudio, indicando aspectos como tipo de investigación, participantes, procedimiento, instrumentos y técnica de análisis de los datos.</p> <ol style="list-style-type: none"> 8. Se describe el tipo de estudio llevado a cabo como, por ejemplo, si fue cuantitativo, cualitativo o mixto; el alcance (descriptivo o descriptivo-correlacional), la lógica de análisis utilizada (deductiva o inductiva), y el tiempo de realización de la investigación (transversal o longitudinal). 9. Se describen los participantes, con elementos tales como: el tamaño de la muestra, el tipo de personas que hicieron parte del estudio y sus datos demográficos, la forma de obtención de la muestra, y los criterios de inclusión y exclusión. 10. Se describe el instrumento o instrumentos empleados, indicando sus autores y los datos de validez y confiabilidad que se posean. Si no se aplicaron instrumentos, entonces se indica la técnica de recolección de los datos empleada en el estudio y la forma cómo se llevó a cabo. 11. Se describen las técnicas utilizadas para el análisis de los datos recolectados (técnicas estadísticas o cualitativas), así como también los criterios éticos aplicados.
IV. Resultados	<p>Es la descripción de los hallazgos más importantes acorde con los principales propósitos del estudio, en forma de texto o mediante tablas o figuras.</p> <ol style="list-style-type: none"> 12. Los resultados se describen de forma sistemática, organizada y sintética, mostrando los aspectos más relevantes y originales del estudio, considerando, en lo posible, el mismo orden de los propósitos establecidos. Se organizan desde lo más general a lo más particular. 13. Se emplean tablas y figuras que ayudan a sintetizar, contextualizar, clarificar o ilustrar los propósitos del estudio. La información contenida en las tablas y figuras no se repite en el texto. 14. Se presentan análisis de los datos de acuerdo con el tipo de estudio; por ejemplo, en los estudios descriptivos cuantitativos es común emplear porcentajes, medias y desviaciones estándar, análisis de cuartiles, diferencias de medias, y análisis de regresión, etc.

Elaboración propia.

CONT. CUADRO 1
Estructura de la escala

V. Discusión	<p>Es el análisis crítico sobre los hallazgos más importantes del estudio, su relevancia científica y sus limitaciones; junto con recomendaciones para aplicaciones, estudios futuros y nuevas líneas de investigación.</p> <p>15. Se presenta una conclusión por cada uno de los propósitos del estudio, con base en los resultados obtenidos, y esta conclusión se analiza con detalle mediante la comparación con estudios similares, que la apoyen o la controviertan, con análisis crítico.</p> <p>16. Se describen las contribuciones más originales o de mayor impacto del estudio, las posibles aplicaciones prácticas de los resultados y también las limitaciones de la investigación (por ejemplo, dificultades metodológicas, deficiencias en el muestreo, problemas con el diseño de la investigación, posibles sesgos, etc.).</p> <p>17. Se brindan recomendaciones para futuros estudios, considerando la experiencia de la investigación llevada a cabo, y teniendo en cuenta las tendencias en el área. En lo posible, esto debe ser con base en argumentos y considerando las propuestas de otros autores.</p>
VI. Referencias	<p>Es la descripción de las fuentes citadas acorde con las normas de la revista.</p> <p>18. Se describen todas las referencias citadas en el texto, acorde con el estilo APA en su última edición, o considerando las normas de la revista en la cual se espera publicar el artículo. Todas las referencias poseen DOI, o, en su defecto, el link de donde se pueden descargar.</p>
VII. Apéndices	<p>Son archivos o información complementaria para comprender el estudio descrito en el artículo, como los instrumentos o la base de datos.</p> <p>19. Se presentan los apéndices cuando el tipo de estudio lo requiere, mediante información extra al final del artículo o mediante archivos complementarios al texto (subidos a la revista o en forma de links a páginas web externas). La información que va en los apéndices no ha sido publicada.</p>
VIII. Estilo y Formato	<p>Consiste en la aplicación del estilo y el formato APA en su última edición, o la aplicación de las normas de la revista.</p> <p>20. El artículo sigue las normas APA en su última edición, o las normas de la revista en la cual se espera publicar. Además, cumple con las normas gramaticales de la lengua española. La redacción es en forma impersonal, atractiva desde el inicio hasta el final, girando todo en torno a los propósitos del estudio. Cada párrafo es argumentativo y se conforma con al menos 7 líneas.</p> <p>21. El formato está acorde con las normas de la revista cuyas consideraciones pueden ser, en términos generales: interlineado a espacio y medio, márgenes de 2.54 cm, letra Times New Roman 12 y numeración continua en la parte superior derecha. En el texto no aparecen los nombres de los autores ni información que los pueda identificar.</p>

Elaboración propia.

Validez de contenido

El siguiente paso fue someter el instrumento a validez de contenido mediante la evaluación cualitativa y cuantitativa de juicio de expertos en los criterios de pertinencia y redacción. Participaron 14 jueces, cuyos datos sociodemográficos aparecen en el cuadro 2. Puede observarse que todos tenían amplia experiencia investigativa y docente, como también alta productividad en publicaciones. Esto demuestra que todos los jueces tenían la experticia requerida para este proceso. Los jueces completaron una escala de validez de contenido elaborada por CIFE (2015), la cual valora la pertinencia y redacción a través de cuatro niveles (1-4), donde 1 es el nivel muy bajo y 5 el nivel muy alto. Luego, se sistematizaron los datos y se obtuvo la V de Aiken (Aiken, 1985), siguiendo la modificación propuesta por Penfield y Giacobbi (2004). El criterio establecido para aceptar los ítems fue que tuvieran un valor de V igual o mayor a 0,80, aunque se aceptaron valores superiores a 0,75 de acuerdo con lo propuesto por otros autores (Bulger y Housner, 2007). Si un ítem presentaba un valor menor a 0,80 pero superior a 0,75 se revisaba y se mejoraba.

CUADRO 2
Datos sociodemográficos de los jueces

N	14
Edad (media \pm desviación estándar)	43 años ($\pm 8,58$)
Género (%)	57% Hombres 43% Mujeres
Nivel educativo	71% Doctorado 29% Maestría
Roles	Coordinadores de posgrado Administrativos Profesores de educación superior Investigadores
Último nivel de estudios (%)	71,43% Doctorado 28,57% Maestría
Áreas de experiencia profesional (%)	71,42% Docencia 14,29% Diseño de instrumentos 14,29% Procesos organizacionales
Años de experiencia profesional (media \pm desviación estándar)	18,21 ($\pm 8,27$)
Años de experiencia docente investigativa (media \pm desviación estándar)	8,42 ($\pm 4,81$)
Artículos publicados en el área (media \pm desviación estándar)	6,85 ($\pm 6,21$)
Libros publicados en el área (media \pm desviación estándar)	1 ($\pm 1,24$)
Capítulos de libro publicados en el área (media \pm desviación estándar)	1,35 ($\pm 1,49$)
Experiencia en el análisis, revisión, diseño y/o validación de instrumentos de investigación (%)	100%

Elaboración propia

Prueba piloto

Una vez la escala obtuvo valores iguales o superiores a 0,80 en la V de Aiken, y se mejoró a partir de las sugerencias de los jueces, se aplicó a un grupo piloto de 6 investigadores de alto nivel, pertenecientes al Sistema Nacional de Investigadores de México. A este sistema se accede por concurso público de méritos, siendo las publicaciones el factor de mayor peso. En el estudio piloto se analizó el grado de comprensión de las instrucciones, los ítems y la forma de respuesta, como también la pertinencia, claridad y grado de satisfacción con el instrumento (cuadro 3). También se les pidió que brindaran sugerencias puntuales para mejorar los indicadores en caso de ser necesario. La aplicación se hizo a través de un formulario en línea con la metodología propuesta por CIFE (2017, 2018).

CUADRO 3
Características sociodemográficas del grupo piloto

N	6
Edad (media \pm desviación estándar)	42 años ($\pm 7,86$)
Género (%)	66,67% Mujeres 33,33% Varones
Último grado de estudios (%)	50% Doctorado 33,34% Posdoctorado 16,66% Maestría
Artículos publicados en revistas indexadas y arbitradas (media \pm desviación estándar)	11,14 ($\pm 6,16$)
Libros publicados por editoriales académicas reconocidas (media \pm desviación estándar)	1,5 ($\pm 1,49$)
Capítulos de libros publicados por editoriales académicas reconocidas (media \pm desviación estándar)	3,85 ($\pm 3,72$)
Ponencias publicadas en memorias o actas de congresos (media \pm desviación estándar)	15 ($\pm 14,17$)

Elaboración propia.

Validez de constructo y confiabilidad

La validez de constructo y confiabilidad de la escala se determinó mediante la aplicación a un grupo de 129 investigadores, también pertenecientes al Sistema Nacional de Investigadores de México. Para ello, se invitó a 2.500 de ellos, pertenecientes al área de ciencias sociales y humanas. Sólo se invitó a quienes tenían un correo público en la base de datos del CONACYT (2017). A todos se les invitó por correo a participar en el estudio y se les pidió su consentimiento informado. Como motivación, se les ofreció una constancia por la aplicación del instrumento a nombre de la universidad del primer coautor de este estudio. No hubo ninguna otra recompensa. Los 129 investigadores que aceptaron participar completaron el instrumento en línea, considerando el último artículo que hubiesen elaborado y publicado en una revista indexada. En el cuadro 4 se describen las características demográficas de este grupo. Se comprueba la alta productividad científica de los investigadores, pues el promedio de artículos publicados por cada investigador fue de 56,2 con una desviación estándar de 85,49. La muestra alcanzada es muy elevada, dado que los investigadores de alto nivel de productividad científica casi no tienen tiempo para responder este tipo de instrumentos.

Para el análisis de validez de constructo, definido como el “grado en que la evidencia y la teoría permiten interpretar las puntuaciones de un test construido bajo un propósito” (Ventura, 2017, p. 1), se empleó el análisis factorial exploratorio (AFE) (Ferrando y Anguiano-Carrasco, 2010; Mavrou, 2015), previa evaluación de la pertinencia de los datos al AFE mediante el índice kmo y el índice de Bartlett, así como una exploración de la matriz de correlación (Costello y Osborne, 2005; Juárez-Hernández, 2018). Si estos coeficientes denotaban pertinencia, se efectuó el AFE mediante el método de extracción de factorización de ejes principales (De Winter y Dodou 2012). Para determinar el número de componentes a retener, se aplicó la regla de Kaiser. Para la significancia de las cargas factoriales, se consideró la propuesta de Rositas-Martínez (2014) con un valor aceptable de 0,5 de acuerdo con el tamaño de la muestra, y si era necesario, se efectuó una rotación de la matriz factorial mediante el algoritmo de mayor conveniencia. Finalmente, la confiabilidad de la escala se determinó mediante el coeficiente Alfa de Cronbach.

CUADRO 4
Características del grupo de investigadores de alto nivel

N	129
Edad (media \pm desviación estándar)	53 años ($\pm 12,67$)
Género (%)	75% Hombres 25% Mujeres
Último grado de estudios (%)	97,68% Doctorado 2,32% Maestría
Años de experiencia profesional (media \pm desviación estándar)	26,54 ($\pm 14,12$)
Experiencia docente investigativa (media \pm desviación estándar)	22,79 años ($\pm 44,58$)
Artículos publicados en el área (media \pm desviación estándar)	56,20 ($\pm 85,49$)
Libros publicados en el área (media \pm desviación estándar)	3,07 ($\pm 5,15$)
Capítulos de libro publicados en el área (media \pm desviación estándar)	10,03 ($\pm 15,73$)
Experiencia en revisión, diseño y validación de instrumentos	86,82% tiene experiencia y 13,18% no
Cargos que desempeñaban durante la aplicación del instrumento	Administrativos Profesores investigadores Investigadores tiempo completo
Áreas de experiencia profesional	Salud, Docencia, Ciencias Sociales, Investigación, Arte, Diseño de instrumentos, Literatura y Matemáticas

Elaboración propia.

3. RESULTADOS

3.1. Validez de contenido

La aplicación de la escala en el grupo de expertos arrojó valores de V de Aiken superiores a 0,80 en la escala de pertinencia, tal y como puede observarse en el cuadro 5. Respecto a la redacción, dos ítems tuvieron valores inferiores a 0,80, pero superiores a 0,75, como fueron los ítems 2 y 12. Debido a que los valores fueron superiores a 0,75, no se eliminaron, sino que se procedió a mejorar su redacción, siguiendo las recomendaciones de los jueces, las cuales se referían a tener mayor claridad y detalle de lo que se pretendía evaluar.

CUADRO 5
Resultados por ítem del coeficiente V de Aiken para los criterios analizados

ÍTEM	PERTINENCIA	REDACCIÓN
1	0,9524	0,8095
2	0,9048	0,7857
3	0,9286	0,8333
4	0,9048	0,8571
5	0,9048	0,8571
6	0,9286	0,8571
7	0,9286	0,9286
8	0,8810	0,8571
9	0,9286	0,9048
10	0,9286	0,9048
11	0,9524	0,8810
12	0,9286	0,7857
13	0,8095	0,8333
14	0,9048	0,9048
15	0,9286	0,9048
16	0,9524	0,8333
17	0,9286	0,9048
18	0,9048	0,9286
19	0,8905	0,9048
20	0,9048	0,8571
21	0,8571	0,8571

Elaboración propia

3.2. Grupo piloto

La aplicación de la escala al grupo piloto arrojó los resultados descritos en el cuadro 6. Se resalta que el tiempo promedio de aplicación del instrumento fue de sólo 12,50 minutos, a pesar de tratarse de una herramienta que evalúa de manera completa un artículo en todos sus componentes y que exige análisis. Además, puede observarse que, en general, la escala es comprensible por parte de los investigadores y los ítems son considerados relevantes (valores superiores al 80% en cada variable, sumando los niveles medio y alto).

CUADRO 6
Valoración del desempeño de la escala en el grupo piloto

CARACTERÍSTICAS	DESEMPEÑO
Modo en que se respondió el instrumento	100% Electrónico
Tiempo promedio en responder la escala	12,50 minutos
Grado de comprensión de los ítems	50% alto/ 33,33% medio/ 16,67% bajo
Grado de relevancia de los ítems	50% alto/ 33,33% medio/ 16,67% bajo

Elaboración propia

3.3. Validez de constructo y confiabilidad

El estudio general de validez de constructo y el análisis de los coeficientes, mostró un valor de KMO: 0,911, X²: 1.578.180; df =210; p= 0,000010, lo cual indica que sí es pertinente realizar el análisis factorial. Los resultados del AFE, de acuerdo con la matriz de comunalidades, permitió identificar valores apropiados para la mayoría de los ítems y de acuerdo con el método de extracción, un único factor posee un autovalor superior a uno que explica más del 64% de la varianza y el 90% de los ítems poseen cargas factoriales representativas (cuadro 7). Dos ítems presentaron cargas factores menores a 0,50 por lo cual se eliminan del instrumento (ítems 2 y 5). La confiabilidad del instrumento, medida por el coeficiente Alfa de Cronbach, fue de 0,937, y muestra que es alta.

CUADRO 7
Comunalidades y carga factorial de los ítems del instrumento

ÍTEM	COMUNALIDAD	CARGA FACTORIAL AL FACTOR 1
1	0,523	0,607
2	0,398	0,420
3	0,682	0,744
4	0,534	0,647
5	0,339	0,483
6	0,548	0,668
7	0,541	0,687
8	0,637	0,774
9	0,682	0,730
10	0,433	0,602
11	0,624	0,681
12	0,523	0,651
13	0,600	0,703
14	0,728	0,780
15	0,612	0,705
16	0,567	0,596
17	0,411	0,587
18	0,620	0,656
19	0,517	0,580
20	0,602	0,685
21	0,697	0,727

Elaboración propia.

4. DISCUSIÓN

Puede afirmarse que la principal contribución del presente estudio fue obtener la Escala de Evaluación de Artículos Descriptivos, con validez de contenido y constructo, como también de confiabilidad. Esto queda demostrado por los valores en la V de Aiken superiores a 0,75 (Bulger, y Housner, 2007; Cicchetti, 1994; Penfield, y Giacobbi, 2004), la buena valoración por el grupo piloto, la prueba de análisis factorial que identificó un único factor acorde con lo esperado (la estructura de un artículo científico) y la alta confiabilidad, superior a 0,90 (Cronbach, 1951). Esto está acorde con las sugerencias y lineamientos básicos sobre cómo debe ser la validación de un instrumento de investigación (Fernández-Cruz, Fernández y Rodríguez, 2018; Páramo, 2017; Soriano, 2014; Tourón, et al., 2018). Hay que agregar, además, que la muestra empleada fue aceptable para el análisis factorial (Costello y Osborne, 2005).

Del instrumento original se eliminan dos ítems: ítem 2, relacionado con los nombres de los autores y el ítem 5 referente a iniciar la introducción con un párrafo que justifique, muestre la relevancia del tema e invite a la lectura del artículo, algo contrario a lo sugerido por Murillo, Martínez-Garrido y Belavi (2017). Esto puede deberse a que falta fortalecer más este aspecto en la forma cómo redactan los artículos los investigadores en la actualidad, a la percepción de que no es algo importante o a la redacción (respecto al ítem 2, el valor de V de Aiken en redacción estuvo por debajo de 0,80, aunque superior a 0,75).

La escala propuesta hace una contribución relevante a la generación de instrumentos para evaluar artículos científicos por cuanto se basa en la aplicación y evaluación por parte de un grupo de científicos destacados, con alto grado de productividad, pertenecientes al Sistema Nacional de Investigadores de México, tal y como quedó demostrado en la descripción de las características de los investigadores. Además, fueron de una misma área (ciencias sociales y humanas) y su número (139 investigadores) fue alto, teniendo en cuenta que se trata de profesionales muy ocupados y que este tipo de actividades no son valoradas como tal en su productividad científica (CONACYT, 2017). De acuerdo con una revisión hecha en diferentes bases de datos como Scopus, Web of Science, Scielo y Redalyc durante diciembre de 2018, por lo menos en México es la primera experiencia de construcción de una escala basada en una muestra de esta naturaleza. Se empleó la fórmula: ("validez" OR "confiabilidad") AND ("artículo científico" OR "artículo académico") en el periodo 2013-2018 (se buscaban instrumentos recientes, de los últimos cinco años).

Un elemento por destacar en la escala diseñada y validada en la presente investigación es que se trata de un instrumento que posibilita a un investigador novel ir autoevaluando un artículo a medida que lo va elaborando, y de esta forma puede lograr que el documento quede con los elementos esenciales que hagan más probable su publicación en una revista indexada, por lo menos de un nivel aceptable, considerando los retos internacionales del conocimiento (COPE, 2018; Elsevier, 2018). Este instrumento también puede ser útil para los pares, con el fin de co-evaluar los artículos de los compañeros, y también podría ser adoptado por algunas revistas cuyos criterios para evaluar artículos científicos son bastante subjetivos.

La Escala de Evaluación de Artículos Descriptivos es semejante a la creada por Rakedzon y Baram, (2017), en el sentido de que se elaboró en forma analítica, pero es distinta porque no se trata de una rúbrica. Además, aquella se diseñó a la par del desarrollo de un curso de escritura académica para estudiantes doctorales de cultura oriental en el área de ingeniería, mientras que la presente se aplicó a investigadores mexicanos con trayectoria de alto nivel en la escritura de investigaciones, procedentes de la cultura latina y perteneciente a las áreas de ciencias sociales y humanidades. Otra diferencia es que el instrumento de Rakedzon y Baram (2017) sólo se hizo mediante el análisis de expertos (validez de contenido), mientras que la presente escala se enmarcó en una aplicación a una muestra importante de investigadores y tuvo en cuenta las sugerencias que hacen editores de revistas prestigiosas indexadas en Scopus (2018), y WOS (2018), las cuales son las dos bases de datos más importantes del mundo.

El presente estudio, comparado con los de Merma, Peña y Peña-Alfaro (2017) y el de Greenberg (2015), son semejantes solo en el sentido de que abordan la escritura científica desde la perspectiva del estilo de la APA, mientras que la presente escala guía y evalúa los aspectos esenciales de contenido y formato, además del estilo antes mencionado, teniendo en cuenta algunos aspectos poco abordados en los instrumentos citados como el énfasis en el abordaje de un problema de investigación, la presentación de los resultados en el mismo orden de los propósitos y la discusión centrada en el análisis crítico de las conclusiones.

Aunque la aplicación de la Escala se hace en poco tiempo (un promedio de 12,50 minutos) y su utilización es muy sencilla y amigable, es muy subjetiva en los niveles, un problema que es común a todas las escalas tipo Likert (Marín-García, y Santandreu, 2015). Por ejemplo, respecto al ítem 6 “Se hace una revisión crítica sobre los principales estudios realizados sobre el problema, considerando los propósitos del estudio, desde lo más general a lo particular”. Además, se basa en citas parafraseadas en la mayoría de los casos, integrando información reciente con información histórica”, la valoración puede ser “nivel muy bajo”, “nivel bajo”, “nivel medio”, “nivel medio alto” o nivel alto”, pero ¿qué elementos concretos hacen que este ítem se cumpla en el nivel medio y no en el nivel medio alto? Concretamente, no hay una respuesta, y los investigadores noveles requieren datos concretos para poder mejorar, por lo cual se recomienda continuar esta línea de investigación y construir una rúbrica analítica que posee más beneficios que la escala (Marín-García, Ramírez, y Atares, 2015; Marín-García y Santandreu, 2015), considerando las contribuciones de la socio-formación (por ejemplo, Ambrosio-Prado, 2018; Tobón et al., 2018). Además, sería recomendable tener como base artículos de alto nivel de calidad, artículos de calidad media y artículos de calidad baja para establecer en qué medida la rúbrica que se elabore ayuda a diferenciarlos.

Entre las conclusiones encontradas en este trabajo se destaca: i) se presenta una escala con los elementos esenciales para contribuir a la evaluación de artículos en investigadores noveles, de ciencias sociales y humanas, con adecuados niveles de validez y confiabilidad, conformada finalmente por 19 ítems (se eliminaron dos a partir del análisis factorial), ii) el instrumento hace una contribución a la evaluación de artículos tomando en cuenta las sugerencias de investigaciones de alto impacto, pertenecientes al Sistema Nacional de Investigadores de México, en el área de las ciencias sociales y humanas. Son pocos los instrumentos validados que tengan este tipo de validación y, por último, iii) se sugiere para futuros estudios elaborar y validar una rúbrica analítica que apoye a los investigadores noveles en la elaboración de un artículo, con descriptores puntuales sobre los niveles de desempeño, algo de lo cual carece la escala validada en el presente estudio.

REFERENCIAS

- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45, 131-142. <https://doi.org/10.1177/0013164485451012>
- Ambrosio-Prado, R. (2018). La socioformación: Un enfoque de cambio educativo. *Revista iberoamericana de educación*, 76(1), 57-82. <https://doi.org/10.35362/rie7612955>
- APA. (2010). *Manual of the American psychological association*. Washington DC: American Psychological Association.
- Bulger, S. M. y Housner, L. D. (2007). Modified Delphi investigation of exercise science in physical education teacher education. *Journal of Teaching in Physical Education* 26(1), 57-80. <https://doi.org/10.1123/jtpe.26.1.57>
- Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessments*, 6(4), 284-290. <https://doi.org/10.1037//1040-3590.6.4.284>
- Costello, A. y Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*, 10(7), 1-9.
- COPE. (2018). *Promoting integrity in research its publication*. Recuperado de <https://publicationethics.org/>
- CIFE. (2015). *Cuestionario de factores sociodemográficos*. Ciudad de México: CIFE.

- CIFE. (2017). *Encuesta de satisfacción sobre el instrumento para el grupo piloto*. Ciudad de México: CIFE.
- CIFE. (2018). *Escala de validez de contenido de CIFE*. Ciudad de México: CIFE.
- CONACYT. (2017). *Tabla de investigadores vigentes 2017 por grado, nivel, género, adscripción, entidad federativa y área de conocimiento*. Recuperado de <https://datos.gob.mx/busca/dataset/sistema-nacional-de-investigadores>
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297- 334. <https://doi.org/10.1007/BF02310555>
- Debnat, J. y Venkatesh, M. G. (2015). Writing and publishing a scientific paper: Facts, myths and realities. *Medical Journal Armed Forces India*, 71, 107-111. <https://doi.org/10.1016/j.mjafi.2015.02.009>
- De Winter, J. C. y Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and simple size. *Journal of Applied Statistics*, 39, 695-710. <https://doi.org/10.1080/02664763.2011.610445>
- Elsevier. (2018). *Publishing with Elsevier: Step-by-step*. Recuperado de <https://www.elsevier.com/authors/journal-authors/submit-your-paper>
- Fernández-Cruz, F. J., Fernández, M. J. y Rodríguez, J. M. (2018). Diseño y validación de un instrumento de medida del perfil de formación docente en tecnologías de la información y comunicación. *Revista Española de Pedagogía*, 36(270), 247-270. <https://doi.org/10.22550/REP76-2-2018-03>
- Ferrando, P. y Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología. *Revista de Psiquiatría y Salud Mental*, 31(1), 18-33.
- González, P. (2018). *Validación de un instrumento de evaluación de las competencias adquiridas en las prácticas clínicas del grado de enfermería de la UCM*. Madrid: Ediciones Universidad Complutense de Madrid.
- Greenberg, K. P. (2015). Rubric use in formative assessment: A detailed behavioral rubric helps students improve scientific writing skills. *Teaching of Psychology* 42(3), 211-217. <https://doi.org/10.1177/0098628315587618>
- Juárez-Hernández, L. G. (2018). *Manual práctico de estadística básica para la investigación*. Mount Dora, FL: Kresearch. <https://doi.org/10.24944/isbn.978-1-945721-24-3>
- LaPlaca, P., Lindgreen, A. y Vanhamme, J. (2017). How to write really good articles for premier academic journals. *Industrial Marketing Management*, 68, 202-209. <https://doi.org/10.1016/j.indmarman.2017.11.014>
- Marín-García y Santandreu. (2015). ¿Qué sabemos sobre el uso de rúbricas en la evaluación de asignaturas universitarias? *Intangible Capital*, 11(1), 118-145. <https://doi.org/10.3926/ic.538>
- Marín-García, J., Ramírez, L. y Atares, L. (2015). Comparing advantages and disadvantages of rating scales, behavior observation scales and paired comparison scales for behavior assessment of competencies in workers. A systematic literature review. *Working Papers on Operations Management*, 6(2), 49-63. <https://doi.org/10.4995/wpom.v6i2.4032>
- Mavrou, I. (2015). Análisis factorial exploratorio: Cuestiones conceptuales y metodológicas. *Revista Nebrija de Lingüística Aplicada*, 19. Recuperado de <https://www.nebrija.com/revista-linguistica/analisis-factorial-exploratorio.html>
- Merma, G., Peña, H. y Peña-Alfaro, S. R. (2017). Design and validation of rubric to assessment the use of American psychological association style in scientific articles. *Journal of New Approaches in Educational Research*, 6(1), 78-86. <https://doi.org/10.7821/naer.2017.1.220>
- Montero, I. y León, O. (2005). Sistema de clasificación del método en los informes de investigación en psicología. *International Journal of Clinical and Health Psychology*, 5(1), 115-127.
- Montero, I. y León, O. (2007). A guide for manning research studies in psychology. *International Journal of Clinical and Health Psychology*, 7(3), 847-862.
- Murillo, F. J. y Martínez-Garrido, C. (2019). Una mirada a la investigación educativa en América Latina a partir de sus artículos. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 17(2), 5-25. <https://doi.org/10.15366/reice2019.17.2.001>

- Murillo, F. J., Martínez-Garrido, C. y Belavi, G. (2017). Sugerencias para escribir un buen artículo científico en educación. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 15(3), 5-34. <https://doi.org/10.15366/reice2017.15.3.001>
- Páramo, P. (2017). *La investigación en ciencias sociales: Técnicas de recolección de información*. Bogotá: Universidad Piloto de Colombia.
- Penfield, R. y Giacobbi, P. (2004). Applying a score confidence interval to Aiken's item content- relevance index. *Measurement in Physical Education and Exercise Science*, 8(4), 213-225. https://doi.org/10.1207/s15327841mpee0804_3
- Pereira, M. G. (2014). The writing of scientific articles. *Brazilian Journal Otorhinolaryngology*, 80, 371-372. <https://doi.org/10.1016/j.bjorl.2014.07.004>
- Rakedzon, T. y Baram, A. (2017). To make a long story short: A rubric for assessing graduate students academic and popular science writing skills. *Assesing Writing*, 32, 28-42. <https://doi.org/10.1016/j.asw.2016.12.004>
- Rositas-Martínez, J. (2014). Los tamaños de muestras en encuestas en las ciencias sociales y su repercusión en la generación del conocimiento. *Innovaciones de Negocios*, 11(22), 235-268.
- Santesteban, O. y Núñez, N. (2017). Cómo escribir un artículo científico por primera vez. *Psiquiatría Biológica*, 24(1), 3-9. <https://doi.org/10.1016/j.psiq.2017.01.004>
- Scopus. (2018). *Support on writing academic papers*. Recuperado de <http://crai.ub.edu/en/crai-services/writing-academic-papers>
- Soriano, A. M. (2014). Diseño y validación de instrumentos de medición. *Diálogos*, 14, 19-40. <https://doi.org/10.5377/dialogos.v0i14.2202>
- Tobón, S., Pimienta-Prieto, J. H., Herrera-Meza, S. R., Juárez-Hernández, L. G. y Hernández- Mosqueda J. S. (2018). Validez y confiabilidad de una rúbrica para evaluar las prácticas pedagógicas en docentes de educación media. *Espacios*, 39(45), 1-30.
- Tourón, J., Martín, D., Navarro-Asencio, E., Pradas, S. e Íñigo, V. (2018). Validación de constructo de un instrumento para medir la competencia digital docente de los profesores. *Revista Española de Pedagogía*, 76(269), 25-54. <https://doi.org/10.22550/REP76-1-2018-02>
- Ventura, J. L. (2017). ¿Validez de constructo o validez basada en el constructo? Comentarios a Soler et al. *Revista de Psiquiatría y Salud Mental*, 10(4), 221-221. <https://doi.org/10.1016/j.rpsm.2017.05.003>
- Vinces, L. (2018). *Diseño y validación de una e-rúbrica para evaluar las competencias clínicas específicas en diagnóstico diferencial en pediatría* (Tesis doctoral). Universidad Casa Grande, Ecuador.
- WOS. (2018). *Research assessment*. Recuperado de <http://wok.mimas.ac.uk/>