

Evaluación de la calidad de las revistas científicas de Psicología publicadas en España: consideraciones al trabajo de Alcaín y Román (2005)

Hugo Carretero-Dios, Macarena de los Santos-Roig y Gualberto Buela-Casal
Universidad de Granada

El objetivo de este trabajo es especificar algunos aspectos metodológicos que deben ser considerados a la hora de llevar a cabo una evaluación de la calidad de las revistas científicas de Psicología publicadas en España. Éstos serán presentados a partir del análisis del estudio de Alcaín y Román (2005), y en el que, a través de una encuesta efectuada al profesorado universitario funcionario, se tenía el objetivo de evaluar las revistas científicas de Psicología en función de su calidad, y obtener los índices de impacto de las previamente mejor valoradas. Las consideraciones efectuadas son agrupadas en los siguientes apartados: a) metodología del estudio y selección de la muestra; b) procedimiento para la valoración de las revistas; y c) corrección de la encuesta y resultados del estudio. Se concluye que el trabajo de Alcaín y Román (2005) tiene serias limitaciones para la evaluación rigurosa de la calidad de las revistas españolas de Psicología.

Scientific quality of journals of psychology published in Spain. The aim of this study is focused on the current topic about the scientific quality of journals of psychology published in Spain. Alcaín's work (Alcaín y Román, 2005), whose aims were to analyze Spanish journals of Psychology in order to determine their scientific quality and to obtain the great quality journals' impact factor indexes, is analyzed in detail and comment. The commentaries are ordered in three structured blocks: a) Methodological comments about sample selection and Alcaín's results; b) Journals' assessment procedures; and c) Alcaín's results limitations and considerations. It is considered that of Alcaín's work results are inappropriate to determine either quality or impact of Spanish journals.

En la actualidad, y con su origen en la institucionalización de la ciencia producida a partir del desarrollo industrial del siglo XIX, existe una importante relación entre autor y producto en el campo de la actividad científica: el autor es evaluado en función del producto. Ocurre, no obstante, que evaluar a corto plazo el producto científico y la productividad de los científicos no es una tarea fácil. El interés sobre esta temática es mayor cada día, y en los últimos años, dentro y fuera de la Psicología, ha aparecido un número considerable de trabajos (Buena-Casal, 2001, 2002, 2003; Buena-Casal, Carretero-Dios y De los Santos-Roig, 2002; Buena-Casal y López, 2005; Fine, 2002; Pelechano, 2002a, 2002b; Pérez-Álvarez, 2002; Polaino, 2002; Gil Roales-Nieto y Luciano, 2002; Sternberg, 2003, etc.), sin que haya un consenso sobre cómo evaluar la actividad investigadora (Musi-Lechuga, Olivas-Ávila, Portillo-Reyes y Villalobos-Galvis, 2005).

Para la evaluación del trabajo que realizan los científicos, un hecho extendido es reducir total o parcialmente, y casi en la mayoría de las ocasiones, la actividad científica a la publicación de trabajos en revistas especializadas. La «calidad» de un científico

es estimada en función del medio donde éste publicó, y a partir de esta estimación se toman multitud de decisiones.

Por todo investigador es sabido que el referente de obligada consulta por los comités y comisiones pertinentes para la valoración de las revistas es la empresa privada *Institute for Scientific Information*, ISI, la cual proporciona el denominado «factor de impacto» de cada una de las revistas incluidas en su *Journal Citation Reports*. El uso ampliamente extendido del factor de impacto de las revistas ha causado críticas dentro de la comunidad científica, llegándolo a calificar de «imperio ISIsta» (Gil Roales-Nieto y Luciano, 2002), y de elemento más de difusión empresarial que de calidad científica (Pelechano, 2002b), aunque igualmente advirtiéndose que quizá el problema principal sea del uso que se da al factor de impacto más que el significado pretendido de éste (Garfield, 2003). Sea como fuere, y sin olvidar el extenso debate sobre los índices bibliométricos de las revistas científicas, y las diferentes propuestas en este sentido (Buena-Casal, Fernández-Hermida y López, 2003), el hecho es que el uso de éstos se está extendiendo entre la comunidad universitaria para evaluaciones diversas.

En España, para la disciplina de Psicología, sólo la revista *Psicothema*, publicada por la Universidad de Oviedo y el Colegio Oficial de Psicólogos de Asturias, está incluida en el *Journal Citation Reports*. Sin embargo, el total de revistas de Psicología definidas como científicas y publicadas en España es mucho mayor. A nivel nacional actualmente existen índices similares al factor de

Fecha recepción: 28/4/05 • Fecha aceptación: 20/1/05

Correspondencia: Hugo Carretero Dios

Facultad de Psicología

Universidad de Granada

18071 Granada (Spain)

E-mail: hugocd@ugr.es

impacto que permiten catalogar a las revistas en función de unos requisitos concretos, como es el caso del *Índice de Impacto de las Revistas Españolas de Ciencias Sociales, IN-RECS* (Grupo de investigación, evaluación de la ciencia y de la comunicación científica, 2004). De igual forma, y con la intención de hallar los índices de impacto de las revistas de Psicología editadas en España, recientemente ha sido realizado el trabajo *Índices de impacto de las revistas españolas de Ciencias Sociales (Biblioteconomía y Documentación, Psicología, Psiquiatría y Urbanística y Ordenación del Territorio) a partir de las revistas mejor valoradas por pares*, Alcain (2003). Éste ha sido llevado a cabo por el Centro de Información y Documentación Científica, CINDOC del CSIC, a través de una subvención de la Dirección General de Universidades, y de él ha aparecido recientemente un resumen en la revista *Psicothema* (Alcain y Román, 2005), donde se muestran los datos obtenidos para las áreas de conocimiento de la Psicología. A partir de ambos trabajos, tanto del original como de su resumen, pueden extraerse sus objetivos principales. El primero de éstos es «someter a la consideración de todos los profesores e investigadores de la disciplina (profesores titulares y catedráticos de Universidad y Escuelas Universitarias) para su evaluación en función de su calidad científica las revistas del área de conocimiento» (Alcain, 2003, p. 3). De igual modo, este objetivo se hace explícito en el resumen comentado, expresándolo en los siguientes términos: «En el presente trabajo se profundiza en la metodología y el análisis de los resultados correspondientes a los indicadores que miden la calidad del contenido de una manera directa (valoración de los pares y citas)» (Alcain y Román, 2005 p. 180). El segundo de los objetivos es analizar «las referencias citadas en las revistas mejor valoradas» (Alcain y Román, 2005, p. 179), con la intención de obtener un índice de impacto de éstas.

La finalidad del trabajo que aquí se presenta, el cual puede clasificarse como de carácter teórico (Montero y León, 2005), es analizar el estudio de Alcain (2003) y su resumen (Alcain y Román, 2005), abordándose lo referente a los datos sobre la disciplina de Psicología y sus áreas académicas. Lo que se pretende con este análisis es proporcionar un marco de reflexión sobre algunos aspectos metodológicos a tener en cuenta a la hora de llevar a cabo estudios sobre la calidad del contenido de las revistas científicas a partir de la opinión del profesorado. Dichos aspectos serán tratados a colación de los elementos que se estiman deben tenerse en cuenta para futuros trabajos con objetivos similares a los del informe aquí analizado.

Metodología del estudio y selección de la muestra

La primera observación sobre la metodología del estudio iría dirigida a que ni en el informe original de Alcain (2003), ni en su resumen (Alcain y Román, 2005), aparece información que es considerada imprescindible cuando se realiza una investigación mediante encuestas (Rojas, Prados y Pérez, 1998). Los datos sobre el tamaño de la población no se presentan, la información sobre el tipo de muestreo y su justificación es insuficiente, algunos datos son contradictorios, el procedimiento está deficientemente explicado, etc. Al hablar de la encuesta, se señala que lo pretendido ha sido «someter a la consideración de todos los profesores e investigadores de la disciplina...» (Alcain, 2003, p. 3), o que «ha sido consultada prácticamente la población total de cada área de conocimiento» (Alcain, 2003, p. 14), o que ha sido enviada a «prácticamente el 100% de los profesores de todas las universidades

españolas» (Alcain y Román, 2005, p. 181). Según esta información, parece que la encuesta va dirigida a todo el censo, aunque posteriormente se indica que el envío no ha sido a todo el censo, sino a *prácticamente* la totalidad de éste. Con respecto a este proceder, es necesario precisar algunas consideraciones a tener en cuenta.

- 1) La primera estaría dirigida a la falta de información referente a lo que significa *prácticamente* la totalidad de la población. Si el objetivo era el censo, la pregunta sería por qué finalmente se ha trabajado con casi la totalidad, y en función de qué criterios se ha seleccionado ésta. En el apartado metodología de Alcain y Román (2005) no aparece epígrafe alguno dedicado a la descripción de la muestra, y en el informe original (Alcain, 2003, pp. 4-5), sólo se indica que para el área de Psicología fue enviada por correo una «encuesta» a un total de 1.574 profesores, con una lista de 181 revistas para evaluar. Ésta es la única información dada con relación a la selección de la muestra.
- 2) Siguiendo con el estudio resumido en Alcain y Román (2005) hay que señalar que las autoras especifican, dentro del apartado resultados, el número de encuestas enviadas y recibidas por área de conocimiento (véase Tabla 1). Como puede observarse, el número de encuestas enviadas por área de conocimiento es desigual. Dado que se intenta encuestar al censo (todos los profesores funcionarios de cada área), es de suponer, puesto que en el informe no aparece información al respecto, que esta desigualdad refleja adecuadamente la de las áreas. Por otro lado, hay que comentar que los porcentajes de encuestas enviadas luego no son mantenidos en los análisis, ya que el porcentaje de las recibidas es distinto del de las enviadas. Si el objetivo era reflejar la situación proporcional de las áreas, éste no se ha conseguido, porque para el análisis de los datos son consideradas todas las encuestas recibidas, sin llegar a efectuar ninguna corrección que mantenga la proporción inicial de las encuestas enviadas. Puede verse en la tabla 1 la disparidad entre el porcentaje muestral conseguido por área y el inicialmente pretendido. Sobre este dato no es realizado comentario alguno.
- 3) Como dato para reflexionar, no habría que dejar de advertir la tasa de participación del profesorado, es decir, la tasa definitiva de respuesta, tanto la total como por áreas, y no perder de vista su influencia (tabla 1). Hay que señalar que en el trabajo no se hace tratamiento alguno de la *no respuesta*. Como es sabido, el problema de la no obtención de

Tabla 1
Tasa de respuesta de las encuestas enviadas

Áreas de conocimiento de Psicología	Encuestas enviadas	Encuestas recibidas	Tasa de respuesta
Metodología de las CC Comportamiento	199 (12,6 %)	68 (13,25%)	34,17%
Personalidad, Evaluación y Tratamiento	340 (21,6 %)	99 (19,30%)	29,11%
Psicología Básica	291 (18,5 %)	135 (26,32%)	46,39%
Psicología Evolutiva y de la Educación	496 (31,51%)	126 (24,56%)	25,40%
Psicología Social	248 (15,75%)	85 (16,57%)	34,27%
Totales	1.574	513	32,60%

* Entre paréntesis se indica el porcentaje inicial de encuestas enviadas y recibidas por área.

respuesta es uno de los principales dentro de la metodología de encuestas (González, 2002). Existen diversas formas de tratar esta no respuesta, al igual que diferentes procedimientos de imputación y estimación de los datos que faltan, mediatizados éstos por los objetivos que se persigan. Aunque las soluciones son difíciles, las alternativas para su tratamiento existen (González, 2002), y no se encuentra comentario alguno sobre este punto ni en Alcaín (2003) ni en Alcaín y Román (2005).

Procedimiento para la valoración de las revistas

Para obtener las respuestas de los profesores universitarios, Alcaín y Román (2005) especifican que se diseña una encuesta conformada por un total de 181 revistas de Psicología. En esta encuesta, a los profesores se les pide que valoren las revistas en función de la calidad de sus contenidos, aunque sólo para aquellas que conozcan, teniendo que dejar en blanco el resto, con la posibilidad de añadir algunas revistas no presentes en la lista. Las opciones de respuesta que se les facilitan son las siguientes: «muy buena, fundamental para la disciplina» (A); «buena, interesante para la disciplina» (B); «de interés general» (C); «sin interés para la disciplina» (D). Se observan algunas deficiencias que deben ser comentadas:

- 1) Las opciones de respuesta son inadecuadas, ya que en realidad se están mezclando dos elementos relacionados pero distintos. A los profesores se les pregunta por la calidad en la opción A y B (Muy buena y buena, respectivamente), pero junto a esto aparece una coletilla (fundamental para la disciplina e interesante para la disciplina). Un profesor puede considerar una revista como muy buena desde, por ejemplo, la aplicación de los supuestos metodológicos, pero a la vez considerar que por los contenidos en los que se centra la revista, demasiado específicos, puede resultar de bajo interés para la disciplina en general. A esto hay que unirle que luego en la opción C y D ya sólo hace referencia al interés de la revista y no a la calidad en términos de buena y mala. Dentro de la construcción de ítems, las opciones de respuesta deben recoger un solo referente, siendo la mezcla de contenidos de respuesta un error a evitar (Osterlind, 1989) en estimaciones unidimensionales.
- 2) Otro de los inconvenientes que tiene el procedimiento de valoración seguido está en que no es una evaluación exter-

na, es decir, los evaluadores son profesores de Psicología interesados en el resultado, por lo que no es difícil que se produzca el efecto de valorar mejor aquellas revistas en las que cada uno publicó, pues el resultado final puede influir posteriormente en la evaluación de los tramos de investigación de los profesores evaluados, en un concurso oposición, etc. Sin embargo, para la estimación de la calidad científica de una revista, quizá el recurrir a juicios de personas directamente afectadas por el resultado de su juicio no es la forma de proceder más adecuada. Cuando se habla de afectados se habla de la totalidad del profesorado, ya que en la actualidad la evaluación del currículo de éstos, sea cual sea la figura docente dentro de la universidad, se centra principalmente en el apartado publicaciones científicas. Además, este problema se acentúa si, tal y como ha ocurrido en el estudio que es analizado, en la muestra se incluyen a algunos de los directores de las revistas, los cuales, sin duda alguna, son parte interesada.

En la tabla 2, tomada de Alcaín (2003), pueden observarse los datos referentes al porcentaje de profesores por área de conocimiento que publican en revistas españolas y que publican en revistas extranjeras. También puede apreciarse la valoración de los profesores sobre las revistas españolas y extranjeras. Como dato de interés hay que indicar que, curiosamente, si se ordenan los resultados de cada área de conocimiento en función del porcentaje de profesores que publican en revistas españolas, podrá apreciarse como este ordenamiento coincide con el que se establecería en función de la calidad asociada a estas revistas (véase Tabla 2). En otras palabras, parece que aquellos que publican más en revistas españolas tienden a valorar como mejores a éstas, resultado que, como mínimo, debería llevar a la reflexión.

- 3) Otro aspecto a considerar tiene que ver con la siguiente pregunta: ¿qué es lo que hay que valorar? Esta pregunta se refiere no sólo a la idea de calidad que se tiene o a si ésta es o no compartida, sino a algo igualmente relevante; *la calidad de qué cosa*. Según se especifica en Alcaín y Román (2005), el objetivo es la calidad del contenido de las revistas científicas. Sin embargo, esto tiene matices que hacen que aunque 100 profesores digan que conocen, por ejemplo, la revista *Psicothema*, no se pueda asegurar que esos 100 profesores han considerado los mismos elementos para realizar un juicio. Como ya ha sido recogido en otros tra-

Tabla 2
Porcentaje de profesores que publican en revistas españolas y extranjeras y de la valoración realizada por éstos sobre su calidad (tomado de Alcaín, 2003)

Área del conocimiento	Porcentaje de profesores que publicaron en revistas		Porcentaje de profesores que consideran mejores las revistas	
	Españolas	Extranjeras	Españolas	Extranjeras
Metodología de CC Comportamiento	80,9 (6)	82,4	39,7 (6)	85,3
Personalidad, Evaluación y Tratamiento	87,9 (2)	63,6	60,6 (2)	76,8
Psicología Básica	83,7 (5)	63,0	43,0 (5)	83,0
Psicología Evolutiva y de la Educación	88,9 (1)	40,5	62,7 (1)	68,3
Psicología Social	85,9 (3)	57,6	54,1 (3)	76,5

* Entre paréntesis y en negrita aparece la posición en que quedan los porcentajes ordenados de mayor a menor.

bajos (Pelechano, 2002a), la calidad científica no es de las revistas, sino de los artículos que en ella son publicados, y que de igual forma presentarán diferencias entre ellos. Cuando a un profesor se le pregunta sobre la calidad de una revista, éste responde en función de una abstracción de lo que conoce de ésta, artículos que ha leído, datos que encontró en esta revista y que le fueron útiles en un momento dado, etc. Dos personas que juzgan una misma revista pueden no centrarse en un mismo estímulo a evaluar, sino en una abstracción de su experiencia con éste. Si lo que interesara fuera evaluar una revista, lo que se tendrían que juzgar serían los trabajos que incluye periódicamente, y realizar una estimación de todos con unos criterios compartidos por las personas que emitan esa estimación. Nótese cómo con esto es enfatizado el hecho de que la evaluación de la calidad de las revistas científicas requiere de estructuración, y no de opinión, de sistematización en cuanto a qué elementos evaluar, bajo qué criterios y a través de qué personas o jueces.

Corrección de la encuesta y resultados del estudio

Las observaciones que se podrían hacer son variadas, sin embargo, las de mayor interés son las siguientes:

- 1) Para la obtención de las respuestas de los profesores, Alcaín y Román (2005) manifiestan que elaboraron una lista conjunta de revistas para el área de Psicobiología y de Psiquiatría, por, según las autoras, la mayor afinidad de la primera a las ciencias médicas y biológicas que a la Psicología. Esto no tiene justificación según la organización de las áreas académicas en España, ni tampoco desde un punto de vista de coherencia o de afinidad entre disciplinas, pues, de ser éste el caso, sin duda alguna el área de Personalidad, Evaluación y Tratamiento Psicológico sería más afín con Psiquiatría que el área de Psicobiología.
- 2) Una vez en posesión de las respuestas de los encuestados, para la obtención de la valoración de cada revista se procede de la siguiente forma. En primer lugar se calcula el porcentaje de profesores que han valorado la revista como «A», «B», «C» o «D» (Alcaín, 2003). Además, se señala qué porcentaje ha juzgado cada revista del total de profesores que responden a la encuesta, tanto por área como para el total de la Psicología. No obstante, habría que decir que en el resumen de Alcaín y Román (2005), los datos referidos a los porcentajes de profesores que han contestado «A», «B», «C» o «D» se obvian y no son presentados. Sin embargo, hay que tener en cuenta que posteriormente estos porcentajes son utilizados y son la base que sustenta la selección de las revistas mejor valoradas. Así, dada su importancia, van a comentarse remitiendo al lector al informe original de Alcaín (2003).

Alcaín y Román (2005) indican el porcentaje de profesores que juzgan una revista. Sin embargo, para presentar las valoraciones sobre éstas (Alcaín, 2003) no se efectúa el cálculo según las personas que han juzgado a una revista, sino según el total de personas que han respondido a la encuesta, aunque no hayan emitido un juicio de una revista determinada. Esto puede ser claramente observado en el informe completo de Alcaín (2003). Por ejemplo, para la sección de Psicología en general, se indica que la revista *Análisis* y

Modificación de Conducta ha sido juzgada por el 62,38% (320) del total de profesores que respondieron a la encuesta (513). Junto a este dato se señala que la valoración de *Análisis* y *Modificación de Conducta* ha sido: A= 16,37%; B= 25,73%; C= 16,18% y D= 4,09%. Como se puede observar, la suma de esos porcentajes da 62,38%, y esto sólo puede ser posible si se divide entre el total de encuestados, no entre el número concreto de profesores que han emitido un juicio sobre esa revista (caso en que la suma de porcentajes debería ser igual a 100). Esto viene a significar que el valor que obtiene la revista en cada una de las opciones de respuesta no es un porcentaje calculado teniendo en cuenta el profesorado que conoce a la revista (320), sino con el total de profesores que respondieron a la encuesta. Se debería haber presentado el porcentaje de valoración sobre las personas que juzgan cada revista, de lo contrario, las revistas menos conocidas siempre obtendrán un porcentaje inferior en las opciones de valoración, al calcular su porcentaje entre los encuestados y no entre los que juzgan la revista.

- 3) Parece ser que posteriormente se intenta evitar el problema anteriormente comentado. Para ello, se realiza el cálculo de dos índices generales de valoración, denominados índice de valoración A (**Iv A**) e índice de valoración A + B (**Iv A + B**). El cálculo de estos índices se realiza, según Alcaín y Román (2005), para ponderar las respuestas en función del número de votos que ha recibido cada revista. En otras palabras, se busca tener en cuenta el número de profesores que ha emitido un juicio para una revista, en función del total de profesores que podía haber emitido ese juicio, que no es otro que los que han respondido a la encuesta para un área de conocimiento determinada. «*Esto permite establecer una comparación interior del área*» (Alcaín y Román, 2005, p. 181). Puede comprobarse que la explicación de cómo son obtenidos los índices **Iv A** e **Iv B** es diferente, según se considere el trabajo original de Alcaín (2003), o bien su resumen (Alcaín y Román, 2005). En concreto, en Alcaín (2003) se especifica que:

Iv A= Valor de «A» multiplicado por el número de votos expresados, dividido entre 100.

Iv A + B= Valor de «A + B» multiplicado por el número de votos expresados, dividido entre 100.

En Alcaín y Román (2005), se especifica que:

Iv A= Valor de «A» multiplicado por el porcentaje de los que valoran, dividido entre 100.

Iv A + B= Valor de «A + B» multiplicado por el porcentaje de los que valoran, dividido entre 100.

Después de revisar los datos del informe original (Alcaín, 2003), y de su resumen (Alcaín y Román, 2005), es posible llegar a la conclusión de que la fórmula aplicada corresponde a la especificada en Alcaín y Román (2005). Esta fórmula, no obstante, resulta errónea. Sirva el siguiente ejemplo sobre la obtención del **Iv A**, para ilustrar lo que se viene comentando, y que podría extrapolarse al **Iv A + B**, ya que se calcula de igual forma pero considerando además los valores de B. La revista *Psicothema*, para el área de Me-

todología de las Ciencias del Comportamiento (Alcain y Román, 2005), obtiene un Iv A igual a 53,42% (véase tabla 2 de Alcain y Román, 2005). A ese valor, y según Alcain y Román (2005), se ha llegado como sigue:

Porcentaje del total de encuestados que valoraron a *Psicothema* con una «A» (55,88%, n= 38), multiplicado por el porcentaje de los que emitieron un juicio sobre *Psicothema* (95,59%, n= 65), dividiendo este resultado entre 100 (Iv A= $(55,88 \times 95,59) / 100 = 53,42$).

Si lo pretendido es tener en cuenta sólo el número de profesores que ha emitido un juicio para una revista, la fórmula aplicada no sólo no es correcta, sino que se debería haber procedido así:

Iv A= Valor de «A» (55,88) dividido por el porcentaje de los que valoran (95,59), multiplicado por 100 (Iv A= $(55,88/95,59) \times 100 = 58,45$).

Nótese que el nuevo valor simplemente responde al cálculo del porcentaje de personas que contestan «A», de entre el total de los que valoran la revista *Psicothema*. Es decir, $38 \times 100 / 65 = 58,45$. En resumen, si se considera lo que Alcain y Román (2005) exponen que quieren conseguir con estos índices, es fácil concluir que las autoras toman un camino demasiado sinuoso para buscar un dato de fácil obtención, y que además lo hacen equivocadamente.

- 4) Es importante tener en cuenta lo dicho hasta ahora en el punto anterior. Primero porque todos los datos cambian radicalmente, y segundo porque luego se obtiene un «índice de impacto» de aquellas revistas que obtienen un Iv A de al menos el 25%, y un Iv A+B de al menos el 75%. Como se puede concluir, las revistas excluidas y que hubieran sido incluidas si el cálculo hubiera sido el correcto, son muchas. No obstante, esta forma de proceder en cuanto a seleccionar las revistas con un Iv A de al menos el 25%, o un Iv A + B del 75%, es igualmente errónea, no sólo por lo inadecuado del cálculo, sino por la falta de coherencia que presenta, tal y como a continuación se comenta.

La obtención del primer índice (Iv A) carece de sentido. Por ejemplo, dada una revista juzgada por 100 profesores con una valoración de: A= 25%; B= 5%; C= 10% y D= 60% (se trataría de una revista de la que 25 profesores dicen que es muy buena, 5 que es buena, 10 de interés general y 60 sin interés) tendría exactamente el mismo nivel en el ranking del Iv A que una revista que recibe A= 25%; B= 65%; C= 5% y D= 5% (una revista en la que sólo 5 dicen que «sin interés» y 90 la sitúan entre A y B).

Como segunda razón, hay que decir que para calcular el segundo índice de valoración, el Iv A+B, se comete el mismo error que en el caso anterior, pues según la fórmula utilizada tendrían el mismo valor, y por tanto el mismo puesto en el ranking, dos revistas cuya suma de A+B sea igual independientemente del valor de A y de B. Por ejemplo, si cien profesores valoran una revista con: 10 A (muy buena), 90 B (buena), 0 C (de interés general) y 0 D (sin interés); ésta tendría el mismo valor que otra revista que clasifican con 90 A, 10 B, 0 C y 0 D; pues en ambos casos la suma de A+B es igual a 100. Sin embargo, es evidente que no es lo mismo que una revista reciba 90 opiniones de muy buena y 10 de buena, que otra que reciba 10 opiniones de muy bue-

na y 90 de buena. Además hay que tener presente que tampoco es igual el valor de C que el de D. Una posible solución en estos casos es ponderar el valor de las respuestas multiplicando cada valor por un factor corrector, por ejemplo: A×15, B×10, C×5 y D×1, y obtener un valor total de valoración de la revista. No obstante, estos porcentajes estarían afectados por los errores ya comentados.

Si lo que se persigue en el estudio es obtener un ordenamiento de las revistas científicas publicadas en España según su calidad, no hay que olvidar que existen otros procedimientos de escalamiento mucho más útiles que el empleado en el estudio analizado, y que se ajustarían más adecuadamente a los objetivos del estudio. En concreto, recurrir a un escalamiento de orden por pasos sucesivos, y donde la lista de revistas que se presentan se tienen que ordenar en función del criterio que se solicite, que en este caso es la calidad, o bien, enfrentar una revista frente a todas las demás, formando pares, donde el evaluador debe elegir sólo una del par en función de su calidad. Este último procedimiento se rechaza si los estímulos son muchos, ya que los pares resultantes son demasiados, pero en su defecto, el ordenamiento de las revistas es una alternativa a considerar, y proporcionaría el *ranking* pretendido, a través de la ordenación más frecuente, o de procedimientos de análisis más complejos. No obstante, esto conllevaría que todos los «evaluadores» estuvieran ante el mismo estímulo (revista), que todos la conocieran, que todos evaluaran una misma cosa, etc.

- 5) A pesar de que el título original del estudio resumido en Alcain y Román (2005) es «Índices de impacto de las revistas españolas...» (Alcain, 2003), no se calcula el factor de impacto de ninguna revista, tal y como está consensuado (Buela-Casal, 2002), y tal y como ya se ha aplicado a algunas revistas españolas (Buela-Casal et al., 2004). El trabajo se limita a ofrecer los datos de citas de algunas revistas recibidas en un subconjunto de revistas (las encuadradas dentro de los índices Iv A e I A+B), pero sin tener en cuenta, por ejemplo, el número de artículos publicados por cada revista, lo cual determina en parte la cantidad de citas de una revista.
- 6) Por último, y antes de finalizar los comentarios sobre el trabajo de Alcain y sus resultados, hay que resaltar que es más que sorprendente que de los más de 500 profesores que responden a la encuesta, existen porcentajes de *no respuesta* más que llamativos, y supuestamente, según las instrucciones de la encuesta, por «no conocer la revista». Así, un 39,09% no emite un juicio sobre *Papeles del Psicólogo*, la revista oficial del Colegio Oficial de Psicólogos y que edita cerca de 40.000 ejemplares de cada número; un 33% tampoco valora a la *Revista de Psicología General y Aplicada*, la revista más antigua de Psicología en España, y un 14% a *Psicothema*, la única revista española de Psicología incluida en el *Journal Citation Reports*. En otras palabras, cuesta creer y entender cómo es posible que 4 de cada 10 profesores de Psicología no conozcan la revista *Papeles del Psicólogo* y 1 de cada 3 tampoco a la *Revista de Psicología General y Aplicada*. En esta dirección, se podría plantear si el diseño del instrumento de evaluación, el cual contaba con un total de 181 revistas, no ha podido ser un factor de influencia en la elevada tasa de no respuesta.

Discusión y conclusiones

Las conclusiones que pueden obtenerse del análisis presentado sobre el trabajo de Alcain (2003) y su resumen (Alcain y Román, 2005) tienen un alcance aplicado bastante amplio. Se han referido una serie de cuestiones metodológicas a tener en cuenta a la hora de evaluar la calidad del contenido de las revistas científicas a través del profesorado, y se han ido apuntando algunos, no todos, de los errores considerados como más importantes, a la vez que se iban proponiendo posibles alternativas a seguir.

La palabra calidad está de moda, y aunque se esté intentando medir desde distintos ámbitos (Gómez y Palací, 2003), y analizando muy exhaustivamente en diferentes contextos de la Psicología (Muñiz y Fernández-Hermida, 2000; Prieto y Muñiz, 2000), no por ello se sabe de qué se está hablando cuando esta palabra es usada asociada a las revistas científicas. Es comprensible que el profesorado exija una consideración de las revistas dónde se publica, e independientemente de que éstas aparezcan en el *Institute for Scientific Information* o no. Pero esta exigencia no puede ir acompañada de premura y falta de estructuración en la puesta en marcha de proyectos. Primero es necesario saber qué es lo que se va a medir, para qué y por qué (Yela, 1990), y después preguntarse por el cómo. Resulta que la calidad científica del contenido de una revista es difícil de definir, e incluso en ciertos momentos puede resultar inadecuado usar criterios cerrados: «*Con la homogeneización de filtros criterios de bondad y adecuación se dificulta el logro de objetivos de multiplicidad, fomento de la creatividad y de la necesaria pluralidad científica. Con la reducción se obtiene a la larga un empobrecimiento y no una mejora de calidad*» (Pelechano, 2002, pp. 334-335). A pesar de esto, y admitiendo la dificultad de la tarea, tampoco conviene aceptar cualquier criterio que sea «adecuado» para un momento o interés personal, y dejar la reflexión en un mero ámbito de adecuación personal de ese criterio, o del resultado obtenido en un estudio concreto. Es importante actuar con rigor en la estimación de los criterios que se planteen, y evaluar minuciosamente los estudios publicados con relación al tema que ocupa este trabajo. La repercusión de esto es muy importante, dado que esos criterios no sólo son usados para evaluar la «calidad» de las revistas. Los resultados de esta evaluación pueden afectar a la valoración del curriculum de los propios científicos, de la productividad científica de las universidades (Buela-Casal, 2005), o de los trabajos sobre efectividad de los tratamientos psicológicos (Frías y Pascual, 2003).

Sin duda, el uso que se haga de la información que se publique en una revista «científica» es también cuestión del «consumidor». Los usos partidistas de los resultados de investigación es un asunto más de ética científica que de calidad. Así, y dado el interés actual sobre la valoración de las revistas científicas, se llama la atención sobre la necesidad de analizar los trabajos dedicados a tal objetivo. La ética científica no es sólo del que hace uso de los datos, sino del que, cómo no, realiza el trabajo y del que le da difusión pública. ¿Quién revisa a los revisores de los trabajos científicos? (Sternberg, 2003). En la introducción se dijo que la actividad científica genera productos de «libre consumo», y ahora, recordando al profesor Vicente Pelechano (2000), se realiza un símil con la idea de producto que se mantiene. Se puede decir que hoy en día se hace un control bastante riguroso de diferentes productos, y baste con poner el ejemplo de los alimentos. Se busca saber qué comemos, y cómo nos puede afectar. La ciencia también es un alimento, cultural, pero en ocasiones su consumo podría provocar

indigestiones y desarreglos graves para la salud. Convendría efectuar quizá controles más rigurosos de los productos que ven la luz en los distintos medios de difusión científica, y no olvidar que algunos trabajos, si son publicados, pueden resultar perjudiciales, y que por el hecho de estar publicados a posteriori no se les puede poner la etiqueta de ciencia: no todo lo que se publica en revistas científicas es ciencia, ni toda la ciencia que se publica en revistas científicas es de igual calibre científico. Así pues, aquí la responsabilidad también es de los revisores y responsables de que un trabajo sea publicado o no.

Ya se sabe que los científicos son «mortales», como no podría ser de otra forma, y como además ha quedado patente en repetidas ocasiones (Benach de Rovira y Tapia, 1995). Sería conveniente no olvidar que esta «mortalidad» es referente a juicios, subjetividad e intereses, y que en el caso de la evaluación de las revistas científicas, los profesores son parte interesada. ¿Resulta adecuado usar como criterio la opinión de los profesores que luego van a ser considerados teniendo en cuenta sus mismas respuestas? Se podría decir que los profesores que participaron en el estudio no sabían el alcance de los resultados, pero esto no viene a anular el enorme peso del contexto sociocultural que lleva a hablar continuamente de evaluación científica, de revistas, de artículos, de comisiones, tribunales, etc. El profesor está demasiado inmerso en el proceso como para utilizar procedimientos de valoración de las revistas científicas que se basen en su totalidad en lo que los profesores «opinan», sin olvidar que sería muy difícil mantener que la calidad es una cuestión de opinión mayoritaria. Así pues, se piensa que la evaluación de las revistas es algo más que urgente, pero también que debería ser hecha por partes no interesadas ni involucradas en los resultados, con realización periódica de evaluaciones, y con edición pública de resultados, baremos, jueces participantes, etc.

Como conclusión final habría que decir que se está de acuerdo en admitir que la evaluación de la calidad de las revistas de Psicología editadas en España fue y sigue siendo un tema de gran interés en todos los ámbitos académicos de la Psicología. Sin embargo, no se puede olvidar que el baremo de uso extendido y generalizado es el proporcionado por el ISI, lo que hace que queden sin consideración casi la totalidad de las revistas de Psicología españolas. El trabajo de Alcain (2003) y su resumen (Alcain y Román, 2005) podría definirse como un intento de evitar esta falta de consideración de la gran mayoría de revistas de Psicología editadas en España. Sin embargo, una vez revisado y analizado, es necesario decir que se trata de un estudio muy ambicioso en cuanto a sus objetivos, pero muy limitado en cuanto a la metodología utilizada y a las conclusiones que de éste pueden extraerse. De este modo, los resultados aportados no deben ser utilizados para evaluar ni calidad, ni difusión, ni nada similar de las revistas tratadas. Las lagunas metodológicas mostradas en el trabajo original de Alcain (2003) y en su resumen, dirigido a presentar los datos de la disciplina de Psicología (Alcain y Román, 2005), podrían contribuir a una pérdida de credibilidad en sistemas alternativos al utilizado por el ISI, a una aplicación de éste en revistas españolas (Buela-Casal et al., 2004) e incluso al de procedimientos alternativos dirigidos, por ejemplo, al cálculo del alcance internacional de estas revistas (Agudelo, Bretón-López y Buela-Casal, 2003). Se reclama que los profesores y gestores de política científica valoren las limitaciones metodológicas del estudio de Alcain y Román (2005) antes de hacer uso de sus resultados.

Referencias

- Agudelo, D., Bretón-López, J. y Buéla-Casal, G. (2003). Análisis bibliométrico de las revistas de Psicología Clínica editadas en castellano. *Psicothema*, 15, 507-516.
- Alcain, M. D. (2003). *Índices de impacto de las revistas españolas de ciencias sociales a partir del análisis de las revistas mejor valoradas por los pares*. www.cindoc.csic.es/info/informecs2003.doc. Recuperado el 15 de mayo de 2004.
- Alcain, M. D. y Román, A. (2005). Hacia una valoración integrada de las revistas españolas de Ciencias Sociales y Humanas: las revistas de Psicología. *Psicothema*, 17, 179-189.
- Benach de Rovira, J. y Tapia, J. (1995). Mitos o realidades: a propósito de la publicación de trabajos científicos. *Mundo Científico*, 154, 124-130.
- Buéla-Casal, G. (2001). La psicología española y su proyección internacional. El problema del criterio: internacional, calidad y castellano y/o inglés. *Papeles del Psicólogo*, 79, 53-57.
- Buéla-Casal, G. (2002). La evaluación de la investigación científica: el criterio de la opinión de la mayoría, el *factor de impacto*, el *factor de prestigio* y «Los Diez Mandamientos» para incrementar las citas. *Análisis y Modificación de Conducta*, 28, 455-475.
- Buéla-Casal, G. (2003). Evaluación de la calidad de los artículos y de las revistas científicas: propuesta del factor impacto ponderado y de un índice de calidad. *Psicothema*, 15, 23-35.
- Buéla-Casal, G. (2005). Situación actual de la productividad científica de las universidades españolas. *International Journal of Clinical and Health Psychology*, 5, 175-190.
- Buéla-Casal, G., Carretero-Dios, H. y De los Santos-Roig, M. (2002). Análisis comparativo de las revistas de Psicología en castellano con factor de impacto. *Psicothema*, 14, 837-852.
- Buéla-Casal, G., Fernández-Hermida, J. R. y López, W. (2003). Iberindex: Índice Iberoamericano de evaluación de publicaciones científicas. *Papeles del Psicólogo*, 17, 41-44.
- Buéla-Casal, G. y López, W. (2005). Evaluación de las revistas iberoamericanas de Psicología: iniciativas y estado actual. *Revista Latinoamericana de Psicología*, 37, 211-217.
- Buéla-Casal, G., Medina, A., Viedma, M. I., Godoy, V., Lozano, S. y Torres, G. (2004). Análisis del factor de impacto de tres revistas españolas de Psicología. *Psicothema*, 16, 680-688.
- Fine, B. J. (2002). Consume science - anywhere. *Observer*, 15, 3.
- Frías, M. D. y Pascual, J. (2003). Psicología clínica basada en pruebas: efecto del tratamiento. *Papeles del Psicólogo*, 85, 11-18.
- Garfield, E. (2003). The meaning og the impact factor. *Revista Internacional de Psicología Clínica y de la Salud / International Journal of Clinical and Health Psychology*, 3, 363-369.
- Gil Roales-Nieto, J. y Luciano, M. C. (2002). A la calidad por *the quantity* (porque la cantidad no vale). Algunas reflexiones sobre los criterios de evaluación de la evaluación de la calidad de la investigación psicológica. *Análisis y Modificación de Conducta*, 28, 431-454.
- González, S. (2002). *El problema de la falta de respuesta: alternativas para su tratamiento en la construcción de estimadores indirectos*. Tesis doctoral no publicada. Departamento de Estadística e Investigación Operativa: Universidad de Granada.
- Gómez, A. y Palací, F.(2003). El constructo de «Orientación a la Calidad Total» y su medida: el cuestionario OCT. *Psicothema*, 15, 433-439.
- Grupo de investigación, evaluación de la ciencia y de la comunicación científica (2004). *Índice de Impacto de las Revistas Españolas de Ciencias Sociales, IN-RECS*. <http://ec3.ugr.es/in-recs/indice-impacto.shtml>. Recuperado el 24 de febrero de 2005.
- Montero, I. y León, O. G. (2005). Sistema de clasificación del método en los informes de investigación en Psicología. *International Journal of Clinical and Health Psychology*, 5, 115-127.
- Muñiz, J. (1998). La medición de lo psicológico. *Psicothema*, 10, 1-21.
- Muñiz, J. y Fernández-Hermida, J. R. (2000). La utilización de los tests en España. *Papeles del Psicólogo*, 76, 41-49.
- Musi-Lechuga, B., Olivás-Ávila, J.A., Portillo-Reyes, V. y Villalobos-Galvis, F. (2005). Producción de los profesores funcionarios de Psicología en España en artículos de revistas con factor de impacto de la Web of Science. *Psicothema*, 17, 539-548.
- Osterlind, S. (1989). *Constructing test items*. Boston: Kluwer Academic Publishers.
- Pelechano, V. (2000). Modelos de personalidad, ciencia de la personalidad y una propuesta. En V. Pelechano (aut.): *Psicología sistemática de la personalidad* (pp. 23-71). Barcelona: Ariel.
- Pelechano, V. (2002a). Presentación del Monográfico 2002: evaluación de la actividad científica. *Análisis y Modificación de Conducta*, 28, 311-316.
- Pelechano, V. (2002b). Valoración de la actividad científica en Psicología. ¿Pseudoproblema, sociologismo o ideologismo? *Análisis y Modificación de Conducta*, 28, 323-362.
- Pérez-Álvarez, M. (2002). La Psicología en el contexto de la cultura española. *Análisis y Modificación de Conducta*, 28, 405-430.
- Polaino, A. (2002). Ciencia, política y política de la investigación. *Análisis y Modificación de Conducta*, 28, 363-390.
- Prieto, G. y Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-75.
- Rojas, A., Fernández, J. y Pérez, C. (1998). *Investigar mediante encuestas: fundamentos teóricos y aspectos prácticos*. Madrid: Síntesis.
- Sternberg, R. J. (2003). There is no place for hostile reviews. *Revista Internacional de Psicología Clínica y de la Salud / International Journal of Clinical and Health Psychology*, 3, 159-161.
- Yela, M. (1990). Evaluar qué y para qué. El problema del criterio. *Papeles del Psicólogo*, 46/47, 50-54.