

ARTÍCULO ORIGINAL

Revisión sobre autoevaluación educativa: evidencia empírica de su implementación a través de la autocalificación sin criterios de evaluación, rúbricas y guiones

Ernesto Panadero-Calderón

ernesto.panadero@oulu.fi

Universidad Autónoma de Madrid
University of Oulu, Finland

Jesús Alonso-Tapia

jesus.alonso@uam.es

Universidad Autónoma de Madrid

RESUMEN: Esta revisión analiza trabajos sobre autoevaluación agrupados en: (1) autocalificación, o autoevaluación sin criterios de evaluación, (2) autoevaluación usando rúbricas y (3) autoevaluación usando guiones. Tras una búsqueda bibliográfica, se realizó un análisis narrativo del contenido de los diferentes estudios seleccionados. Con respecto a la autoevaluación sin criterios de evaluación destaca la falta de uniformidad en los resultados obtenidos por los distintos estudios aunque, en general muestran que la misma no garantiza un aprendizaje profundo de los alumnos. El uso de rúbricas y guiones presenta mejores efectos en el aprendizaje pero quedan aspectos por investigar: los estudios con rúbricas tienen diseños empíricos débiles y los estudios con guiones no han sido realizados en contextos reales. Las conclusiones indican que ofrece más garantías el uso de rúbricas y guiones. Se señalan líneas futuras de investigación.

PALABRAS CLAVE: Autoevaluación, Autorregulación, Evaluación Formativa, Rúbricas, Guiones de Evaluación, Indicios, Pistas, Auto-Calificación.

Self-assessment effects review: empirical evidence about its implementation using self-grading, rubrics and scripts

ABSTRACT: This review analyzes studies on the effects of self-assessment. These studies are grouped as: (1) self-grading, or interventions to promote self-assessment without assessment criteria, (2) use of rubrics, and (3) use of scripts. A literature search was conducted followed by a narrative content analysis of the selected studies. Regarding self-assessment without assessment criteria, the results are heterogeneous, though, in general, they show that these interventions do not warrant an improvement in the students' learning. The use of rubrics and scripts presents more promising effects, but there are still important aspects to explore: studies on rubrics tend to rely on weak research designs, while studies

on scripts, in general, have not been implemented in real classrooms settings. The main conclusion is that rubrics and scripts offer better conditions for self-assessment. Future lines of research are discussed.

KEY WORDS: Self-Assessment, Self-Regulation, Formative Assessment, Rubrics, Scripts, Prompts, Crues, Self-Grading, Self-Evaluation

Fecha de recepción 08/04/2013 · Fecha de aceptación 09/10/2013
Dirección de contacto:
Ernesto Panadero Calderón
Department of Educational Sciences and Teacher Education,
Learning and Educational Technology Research Unit (LET)
University of Oulu (Finland)
PO BOX 2000, Fin-90014

Agradecimientos

Esta investigación fue llevada a cabo a través de una beca del Ministerio de Educación (España) a Ernesto Panadero (Ref. SEJ2005-00994).

1. INTRODUCCIÓN

La autoevaluación es una práctica habitual en las aulas, como los propios profesores informan (Noonan y Duncan, 2005). Sin embargo, a pesar de estar tan extendida, muchos profesores declaran que no siempre tienen clara su utilidad pedagógica ni cómo fomentarla adecuadamente entre sus alumnos. Este hecho, que a primera vista podría parecer exclusivamente una muestra de desvinculación entre la investigación y la práctica, dado que la autoevaluación ha sido ampliamente estudiada (Boud y Falchikov, 1989; Dochy, Segers, y Sluijsmans, 1999; McDonald y Boud, 2003; Tan, 2012; Taras, 2008), demuestra que quedan aspectos pedagógicos por explorar y trasladar a los entornos reales de enseñanza máxime si, como se va a argumentar en el presente artículo, hay una fragmentación tanto teórica como práctica sobre cómo definir la autoevaluación y cómo fomentarla en el aula.

En primer lugar, es necesario justificar la relevancia de la autoevaluación para el aprendizaje de nuestros alumnos. Hay dos corrientes de investigación muy sólidas que hacen énfasis en la importancia de la autorregulación. La primera proviene de las teorías sobre autorregulación que es “*el control*

que el sujeto realiza sobre sus pensamientos, acciones, emociones y motivación a través de estrategias personales para alcanzar los objetivos que ha establecido” (Zimmerman, 2000, página 14). En esta corriente la autoevaluación es entendida como un proceso que el alumno realiza para autorregular su aprendizaje, enfocado en la detección de aciertos y errores (Winne y Hadwin, 1998; Zimmerman, 2000). La segunda proviene de la evaluación formativa (*formative assessment*) desde la cual se defiende que la autoevaluación es una estrategia de aprendizaje que los profesores deben fomentar, presentándola como estrategia de instrucción (Black y Wiliam, 1998; Nicol y McFarlane-Dick, 2006). Además, la relevancia de la autoevaluación para el aprendizaje se ha puesto de manifiesto, en segundo lugar, en la gran cantidad de evidencia empírica sobre su utilidad y eficacia acumulada en las últimas tres décadas (Boud y Falchikov, 1989; Dignath y Büttner, 2008; Dochy et al., 1999; Falchikov y Boud, 1989; Tan, 2012; Ward, Gruppen, y Regehr, 2002).

A pesar del peso que ha tenido la autoevaluación en la psicología educativa hay aspectos que necesitan más investigación. Durante los últimos años se ha desarrollado el uso de determinados apoyos de instrucción y didácticos a la implementación de la autoevaluación en el aula que no han sido incluidos en las revisiones de autoevaluación ya realizadas (Alonso-Tapia y Panadero, 2010). El objetivo del presente trabajo es precisamente revisar los hallazgos de estas nuevas corrientes de investigación y compararlos con la línea tradicional de estudios sobre autoevaluación.

1.1. Definición de autoevaluación y la importancia de los criterios de evaluación

En este trabajo la autoevaluación se define como: “la valoración cualitativa del proceso de aprendizaje y del producto final obtenido a partir de unos criterios de evaluación”. Esta definición incluye no sólo la valoración del resultado final, sino también del proceso a través del cual se llega a éste, pues ciertamente los alumnos también pueden autoevaluarse durante la realización de una actividad y no sólo al acabarla (Winne y Hadwin, 1998). Además, nuestra definición incluye un aspecto clave: los criterios de evaluación. Estos criterios son las directrices específicas bajo las que se evaluará el trabajo tanto por parte del profesor como del alumno (Goodrich, 1996). Los criterios pueden establecerse de tres formas. Primero, *de forma externa*, situación en la que es el profesor quien los establece; segundo, *de forma mixta, situación en la que* el profesor los propone, discute y negocia con los alumnos, que los aceptan como propios tras la negociación; y, tercero, pueden establecerse internamente por el propio alumno, cuando reflexiona sobre las características que debe tener su trabajo y establece qué debe conseguir y cómo. El grado de autonomía de estas tres formas de establecer los criterios es diferente: cuanto mayor es la sensación de autonomía en la elección de los criterios, mayor motivación habrá por alcanzarlos (Deci y Ryan, 1985).

La importancia de estos criterios para una correcta autoevaluación es crucial pues, haciéndose explícitos, los alumnos están más seguros de lo que deben hacer y cómo evaluar si lo están haciendo correctamente o no, mejorando el aprendizaje (Andrade, 2010; Andrade y Valtcheva, 2009; Boekaerts y Corno, 2005). De hecho, en el listado de condiciones que fomentan la autoevaluación realizado por Goodrich (1996) se puede leer, entre otra serie de acciones: “Tener acceso a los criterios en los que se basa la evaluación”. Precisamente hay un trabajo previo en el que se propone una clasificación de las diferentes ayudas a la autoevaluación basada en la presencia y forma de presentación de los criterios de evaluación (Alonso-Tapia y Panadero, 2010). Dada la relevancia de estos criterios, en el presente

trabajo haremos una revisión atendiendo a la propuesta de estos autores y buscando la evidencia empírica que pueda respaldar cuál de las tres ayudas que ellos clasificaron fomenta mejores entornos de aprendizaje. Estas tres categorías son¹:

1. Entrenamiento en *autocalificación*, o autoevaluación sin entrega de criterios.
2. Entrenamiento en autoevaluación mediante el uso de *rúbricas*.
3. Entrenamiento en autoevaluación mediante el uso de *guiones*.

En suma, el objetivo de esta investigación es analizar la evidencia empírica sobre estas tres formas de implementar la autoevaluación en el aula para explorar las siguientes preguntas de investigación: ¿presenta mejoras el uso de un tipo de ayuda frente a las otras? ¿cómo se han utilizado este tipo de ayudas y cuán concluyentes resultan sus resultados?

2. MÉTODO

2.1. Búsqueda de palabras claves y bases de datos utilizadas

Usando las bases de datos PsycINFO y ERIC se realizó una búsqueda sin limitaciones en el año de publicación. Se emplearon las siguientes combinaciones de palabras restringidas a su utilización en títulos y resúmenes: *self-assessment/self-evaluation, rubric/s, script/s-prompt/s-cue/s*. Para la categoría de autoevaluación sin criterios de evaluación se encontró un número muy elevado de artículos –más de 500– por lo que se limitó la búsqueda a las revisiones localizando 4. En la categoría de rúbricas se encontraron 52 artículos. Tras leer los resúmenes se seleccionaron 18. En la categoría de guiones de autoevaluación se encontraron 83 artículos seleccionando finalmente 25. Las razones para rechazar artículos fueron: (a) diferentes definiciones de lo que era una rúbrica o guión, (b) no habían pasado por revisión entre pares, (c) no evaluaban mejoras en el aprendizaje de los participantes. Posteriormente, usando el método denominado “bola de nieve” (Cooper,

2010) se localizaron más artículos que se incorporaron al estudio.

2.2. Criterios de inclusión

Se evaluaron artículos de todas las disciplinas basando la decisión de incluirlos exclusivamente en su calidad y el cumplimiento de los dos siguientes criterios. Primero, tenían que haber pasado un proceso de revisión entre pares y estar publicados. Segundo, tenían que presentar datos empíricos sobre autoevaluación para poder aportar información a esta revisión. De esta forma los artículos teóricos no fueron considerados para formar parte de los resultados empíricos de este estudio.

2.3. Análisis

El procedimiento de revisión de los estudios seleccionados fue de carácter cualitativo dado que el propósito de este trabajo es identificar el mayor rango posible de efectos producidos por la implementación de las tres formas de autoevaluación. Por ello, no se analizó el tamaño de estos efectos usando técnicas cuantitativas (ejemplo meta-análisis) que, además, habría resultado inviable debido a la gran heterogeneidad de los estudios impidiendo este tipo de análisis estadísticos. Por este motivo el método usado fue el análisis narrativo del contenido (Dochy, 2006).

En un primer paso se leyeron los artículos para determinar la información relevante que contenían y si cumplían los criterios de inclusión. A continuación se extrajo la información que se ha recogido en las tablas que se presentarán en el apartado resultados, menos para los estudios de la primera categoría pues, como se ha comentado, se recurrió a revisiones directamente. En el tercer paso se compararon los diferentes aspectos de los estudios extrayendo conclusiones sobre los mismos. La confección de las tablas de rúbricas y guiones se realizó de forma independiente por los autores, comparándolas después y discutiendo los aspectos no coincidentes y volviendo a los artículos originales a buscar la información en caso de desacuerdo.

3. RESULTADOS

Como ya se ha expuesto, el criterio que se siguió en este trabajo para categorizar los diferentes entrenamientos en autoevaluación ha sido la presencia y tipo de presentación de los criterios de evaluación en la forma de implementar la autoevaluación (Alonso-Tapia y Panadero, 2010).

3.1. Autoevaluación sin entrega de criterios de evaluación o autocalificación

Este procedimiento consiste en que el alumno evalúe y/o califique su trabajo al finalizarlo sin que se le haya hecho entrega de los criterios de evaluación. Históricamente se ha considerado “autoevaluación” a la “autocalificación” (*self-grading*) –ponerse una nota–, asimilación conceptual que en la actualidad se pone en entredicho (Andrade y Valtcheva, 2009; Boud, 1995). Se considera que el mero hecho de ponerse una nota no garantiza que el alumno haya considerado los criterios de evaluación y, por lo tanto, que se autoevalúe de forma adecuada, así como que se produzcan aprendizaje y autorregulación, por lo que en los estudios señalados habría sido más preciso el uso del término “autocalificación”. De hecho, la evidencia empírica encontrada señala que la exactitud de las auto-calificaciones es baja cuando no se hace entrega de un instrumento que ayude a comprender los criterios de evaluación (Andrade y Valtcheva, 2009; Boud y Falchikov, 1989; Falchikov y Boud, 1989; Tan, 2012).

El estudio más importante sobre este procedimiento es el meta-análisis realizado por Falchikov y Boud (1989), en el que analizaron 57 estudios experimentales, la mayoría centrados en la autocalificación. De acuerdo con los resultados obtenidos, la autocalificación presenta varios problemas. Si la tarea es novedosa, los alumnos con experiencia en autoevaluación en otras tareas no son más precisos en sus auto-calificaciones que alumnos sin experiencia. En otras palabras, los alumnos con mayor experiencia en autoevaluación tienen problemas para transferir esta habilidad a nuevas tareas. Sin embargo, sí se encontró que alumnos

con más experiencia en una tarea se autocalificaban de forma más precisa que los que eran novatos.

Otro resultado fundamental del trabajo de Falchikov y Boud (1989) es que, en la mayoría de estudios revisados, el éxito en la autoevaluación se entendía como la aproximación de la autocalificación a la nota del profesor y a la calificación recibida de los compañeros. Por lo tanto, no se evaluaba directamente el aprendizaje, sino la correlación entre la exactitud de la calificación del profesor y la del alumno. Actuar de este modo presenta dos problemas, primero, que no hay evaluación externa de la mejora y, segundo, que los acuerdos entre las calificaciones de los alumnos y los profesores varían significativamente. De hecho esa es una de las mayores críticas que ha recibido la autocalificación: los alumnos que destacan tienden a calificarse por debajo de lo que su trabajo merece mientras que los alumnos con problemas tienden a inflar su nota (Falchikov y Boud, 1989; Kruger y Dunning, 1999; Landrum, 1999; Lejk y Wyvill, 2001; Sadler y Good, 2006).

Por estos motivos la autocalificación fomenta poco la autorregulación pues es una aproximación superficial que no incluye una reflexión profunda al no estar basada en los criterios de evaluación (Andrade y Valtcheva, 2009). Así, los alumnos no saben cómo valorar su tarea para que se ajuste a los criterios con los que el profesor evaluará su trabajo. De esta forma un alumno puede evaluar su trabajo atendiendo a lo que él piensa que es el modelo, pero nada garantiza que éste sea correcto. En caso de que sea erróneo, el alumno estará, por un lado, evaluándose por encima o por debajo de la calidad de su trabajo y, por otro, no estará aprendiendo de los errores cometidos. Por esta razón, es imprescindible que la autoevaluación se fomente proporcionando a los alumnos modelos expertos que incluyan los criterios de evaluación.

Tras la revisión de Falchikov y Boud (1989) se produjo un descenso muy acusado de los artículos que estudiaban la autoevaluación

sólo como autocalificación. A partir de entonces empezó la otra parte de la investigación englobada en esta categoría, que no se centra exclusivamente en la calificación del alumno, sino en que éste autoevalúe el contenido de lo realizado, lo cual sí es beneficioso para su aprendizaje (Dochy et al., 1999; Taras, 2010; Ward et al., 2002). Sin embargo, el problema es que este tipo de investigación no deja claro, primero, si se ofrece a los alumnos los criterios de evaluación y, segundo, en caso de que estos se entreguen, no especifica cómo se hace. Pedir a un alumno que se autoevalúe, aun haciendo hincapié en que no se trata de ponerse una nota, tiene la ventaja de que el alumno se centra en la evaluación de la tarea pero no ofrece las condiciones óptimas para que el alumno lo haga de forma adecuada porque, ¿qué criterios utiliza para autoevaluarse? ¿Son estos correctos?

Finalmente, el trabajo que ha puesto en entredicho todos los trabajos previos ha sido realizado recientemente por Tan (2012). Tan ha encontrado errores en las revisiones posteriores a las realizadas por Falchikov y Boud (1989) llegando a conclusiones erróneas sobre la eficacia de la autoevaluación fomentada sin criterios de evaluación. Al mismo tiempo Tan (2012) mantiene que hay tantos estudios que han encontrado que los alumnos son exactos al autoevaluarse como que no, por lo que concluye que es necesario dotar a los alumnos de ayudas más específicas y estructuradas.

En suma, el mayor problema con esta línea de investigación es que no se hace entrega de los criterios de evaluación. Esto genera, por un lado, que se confunda autoevaluación con autocalificación y, por otro, que no se garantice una correcta autoevaluación por falta de criterios y, en consecuencia, tampoco se garantiza la autorregulación ni el aprendizaje. Debido a ello, en los últimos años ha habido un aumento de intervenciones en autoevaluación usando rúbricas y guiones.

3.2. Autoevaluación a través del uso de rúbricas

Una rúbrica es un instrumento de autoevaluación y calificación que contiene los criterios de evaluación y se compone de

tres partes: un baremo para la autocalificación, un gradiente de la calidad del trabajo y una explicación breve de en qué consiste el estándar para esa categoría (Tabla 1).

Puntuación Criterio de calidad	4	3	2	1
Conceptos	Contiene todos los conceptos importantes y todos los secundarios relevantes	Contiene los principales y algunos secundarios, pero faltan algunos secundarios	Contiene los conceptos principales pero no los secundarios	Falta/n concepto/s principales del tema
Jerarquización	Tiene una organización correcta y completa, y el mapa la transmite adecuadamente	La organización es correcta pero incompleta: faltan niveles o elementos dentro de un nivel	La organización es incorrecta pero completa: hay conceptos mal situados; aparecen en un nivel distinto del que les corresponde	La organización es incorrecta e incompleta: faltan niveles o elementos dentro de un nivel y otros están mal situados
Relaciones entre conceptos de diferente nivel jerárquico	RELACIONES Son correctas: las líneas unen los conceptos que deben unir NEXOS Están explícitos y ayudan a entender mejor las relaciones	RELACIONES Son correctas pero incompletas: faltan líneas que deberían estar presentes NEXOS Incompletos: Sólo se explicitan algunos, pero correctamente	RELACIONES Son parcialmente incorrectas: algunas líneas unen conceptos que no deben NEXOS Incorrectos: Están todos, pero algunos no son correctos	RELACIONES Son incorrectas en su mayoría o inexistentes NEXOS Incompletos e incorrectos.
Relaciones entre conceptos de distintas columnas	Están presentes y añaden información importante	Sólo hay alguno	No hay	No hay
Simplicidad y facilidad de uso	Visualmente es sencillo y claro. Contiene ejemplos	Algunas líneas de relación no se entienden. Contiene sólo algún ejemplo	El número de conexiones es excesivo y hace difícil su uso. No contiene ejemplos	No se entienden bien las relaciones ni la jerarquía. No contiene ejemplos

Tabla 1. Rúbrica para evaluar el procedimiento de construcción de mapas conceptuales. Autores: Ernesto Panadero, Jesús Alonso-Tapia y Juan Antonio Huertas

Durante su uso, los estudiantes comparan su trabajo con los estándares que aparecen en la rúbrica y buscan en el baremo su calificación. Es una ayuda a la autoevaluación que está en crecimiento, con una amplia penetración en Norteamérica (Reddy, 2007) y con un uso creciente en Europa (Simon y Forgette-Giroux, 2001).

Las rúbricas pueden ser de dos tipos: analíticas –se puntúa criterio por criterio– u holísticas –también hay diferentes criterios pero se puntúa con una calificación global–.

Se supone que si los alumnos utilizan las rúbricas aprenden a autoevaluar su trabajo en la fase final, la de auto-reflexión, contrastando su trabajo con los criterios de evaluación que aplicará el profesor. Pero también es interesante,

y de hecho se recomienda, fomentar que los alumnos utilicen la rúbrica durante la planificación de la actividad porque podrán establecer objetivos de trabajo más acordes a la evaluación posterior, por lo que se recomienda hacer la entrega de la rúbrica al inicio de la actividad (Andrade y Valtcheva, 2009).

La pregunta, sin embargo, es: ¿son adecuadas las rúbricas para facilitar la autorregulación y el aprendizaje? Teóricamente, la rúbrica resulta un instrumento fundamental porque explícita y ayuda a entender el criterio de evaluación (Goodrich Andrade, 2001), puede presentar modelos expertos -depende de la calidad con la que esté confeccionada- y ofrece los estándares de calidad (Goodrich Andrade, 1996). Pero, ¿qué dice la evidencia empírica? En la Tabla 2 se recogen los principales estudios empíricos. No obstante, antes de pasar a presentar estos estudios en detalle vamos a comentar los resultados de las dos revisiones sobre rúbricas que hay publicadas. Jonsson y Svingby (2007) revisaron el estado del arte sobre las rúbricas centrados no en los efectos sobre el aprendizaje si no en su capacidad para aumentar la fiabilidad de la evaluación. En sus conclusiones destacan tres:

- 1) La fiabilidad de las calificaciones puede aumentarse con el uso de rúbricas, especialmente si son analíticas, específicas para la tarea y presentan ejemplos o prácticas.
- 2) Las rúbricas no garantizan una autoevaluación correcta *per se* sin embargo, la evaluación puede mejorar usando rúbricas en un marco de evaluación comprensivo.
- 3) Las rúbricas parecen tener potencial para fomentar el aprendizaje y mejorar la instrucción.

Los autores de la segunda revisión sobre rúbricas, Panadero y Jonsson (2013), sí analizaron los efectos de éstas en el aprendizaje llegando a las siguientes conclusiones. El uso de las rúbricas puede tener los siguientes efectos: aumentar el rendimiento académico y la

autorregulación, aumentar la transparencia en las calificaciones, reducir la ansiedad de los alumnos y aumentar la auto-eficacia. Pero un aspecto fundamental para que esto ocurra es que el uso de rúbricas se combine con actividades metacognitivas, es decir, con la promoción de la autoevaluación.

¿Cuáles son las conclusiones que podemos sacar de los estudios analizados en esta revisión (Tabla 2)?

Se ha demostrado que las rúbricas ayudan a comprender el criterio de evaluación (Andrade y Du, 2005), que los estudiantes perciben positivamente su uso y con ellas reducen su nivel de ansiedad (Andrade y Du, 2005), que a menudo, no siempre, contribuyen a que aumente la percepción de autoeficacia (Andrade, Wang, Du, y Akawi, 2009) y, principalmente, que mejoran los resultados académicos y la ejecución (Andrade, Du, y Mycek, 2010; Andrade, Du, y Wang, 2008; Goodrich Andrade, 2001; Hafner y Hafner, 2003; McCormick, Dooley, Lindner, y Cummins,

2007). Estos resultados indican que las rúbricas son importantes porque ayudan al alumno a autoevaluar su trabajo ya que, cuando se tiene un modelo que seguir y se conocen los objetivos principales de una tarea, es más sencillo aprender. Por ello, parece lógico pensar que las rúbricas favorecen la autoevaluación pues poseen dos características fundamentales para que esta última funcione: los criterios de evaluación y los estándares de calidad.

Otro de los aspectos positivos de las rúbricas es que ayudan al profesor a clarificar sus propios criterios de evaluación (Schafer, Swanson, Bené, y Newberry, 2001). El hecho de que el profesor tenga claro cómo va a evaluar y tenga conocimientos sobre el uso de rúbricas con ese fin favorece el aprendizaje de los alumnos.

Sin embargo, aunque las rúbricas han demostrado su utilidad para facilitar la autoevaluación y a, través de ella, el aprendizaje, también presentan varias debilidades (Tabla 3).

Estudio	Muestra	Tarea	Tipo ayuda	Procedimiento	Resultados
Goodrich Andrade (2001)	242 estudiantes 8º grado. Edad no especificada.	Redacción y escritura	Rúbrica para autoevaluar la redacción	Los participantes realizaron tres redacciones separadas en un mes. La 1ª era persuasiva, la 2ª autobiográfica y la 3ª ficción. El experimental recibió una rúbrica específica antes de cada redacción. Tres semanas después los participantes contestaron de forma narrativa a una pregunta.	Entregar una rúbrica no implica <i>per se</i> que aumente el conocimiento de los alumnos sobre el criterio. Este conocimiento implica una comprensión de la rúbrica y su uso por lo que se recomienda que los alumnos participen en su creación.
Schafer et al. (2001)	71 parejas de profesores de secundaria. Edad no especificada.	Algebra, Biología, Inglés, "Gobierno"	Las rúbricas eran específicas para cada asignatura	Seleccionaron parejas de profesores de la misma materia y centro. Todos los profesores acudieron a una sesión de entrenamiento general. Una segunda sesión un miembro de cada pareja recibía entrenamiento específico en rúbricas. Se les entregaba una y se les recomendaba que estuviese en su planificación, utilizándola para dar las instrucciones a los alumnos. Se trabajó en las cuatro áreas. Lo que no queda claro es si controlaron el uso de la rúbrica en clase. Después se evaluó el aprendizaje con dos exámenes diferentes por cada área.	Los datos se analizaron a través de un meta-análisis considerando los datos generados por cada profesor como una investigación independiente. El efecto del uso de la rúbrica sólo fue significativo en Álgebra (para los dos exámenes) y Biología (para uno de los exámenes). El uso de rúbricas es también positivo para el profesor pues le ayuda a tener claros los objetivos y criterios de evaluación.
Andrade y Boulay (2003)	397 estudiantes. 183 alumnos de 7º grado, y 214 de 8º. 191 chicos, 201 chicas.	Escritura	Rúbrica sobre redacción	Los dos grupos recibieron la rúbrica y escribieron dos redacciones. Al grupo de línea base sólo se le entregó la rúbrica, mientras que al grupo de intervención recibió entrenamiento durante dos sesiones para aprender a autoevaluarse usando la rúbrica.	No hubo efecto significativo del entrenamiento recibido por el grupo de tratamiento, pero la tendencia le favorecía. No hubo diferencias basadas en género, pero hubo un efecto significativo de la interacción género-tratamiento: las chicas en el grupo tratamiento obtuvieron mejores notas que las del grupo control.
Hafner y Hafner (2003)	107 estudiantes universitarios de Biología. Edad no especificada.	Presentación Biología	Rúbrica general para presentación en la asignatura de Biología	La rúbrica se utilizó durante tres años consecutivos en el curso de Biología que impartía uno de los autores. Se entregaba al principio para que planificaran una presentación de un trabajo. Se les recomendaba su uso. A la hora de evaluar se realizaba en una plantilla basada en la rúbrica. Evaluaban el profesor, los compañeros y el grupo que presentaban.	Importante: las evaluaciones de los alumnos por parte del profesor las hizo uno de los autores y no se contrastaron con un observador más objetivo. Según el investigador hubo un mejor aprendizaje pues los trabajos eran de mejor calidad que otros años, especialmente los últimos de cada curso pues aprendían de las exposiciones anteriores. Las evaluaciones de los compañeros eran parecidas a las del profesor sobre todo en las notas altas. Las auto calificaciones fueron extremadamente parecidas a las que daba el profesor. Los alumnos percibieron que lo habían hecho mejor con la rúbrica.
Brown et al. (2004)	No resulta claro si usaron la rúbrica para la intervención y, en caso de que lo hicieran, hubo entrenamiento metacognitivo y modelado, así que los efectos observados no se podrían argumentar como causados por la rúbrica.				
Andrade y Du (2005)	14 estudiantes de Magisterio	General para la asignatura	Entrega a todos los estudiantes del guión al principio del curso	Los alumnos trabajaron con una de las investigadoras durante un curso con rúbricas. Al finalizar se les entrevistó en grupo.	Las percepciones de los alumnos del uso de la rúbrica fueron que les ayudaba a determinar las expectativas para la tarea, planificar, revisar su trabajo, guiarles y reflexionar sobre la retroalimentación, tener mejores calificaciones y más justas, mejoras en la calidad del trabajo y menos ansiedad.
Green y Bowser (2006)	Sin datos sobre muestra ni edad. Sin control.	Realizar una revisión literaria.	Rúbrica sobre cómo realizar una revisión literaria	Entrenaron a los alumnos a realizar una revisión literaria.	Es un trabajo descriptivo en el que desarrollan un modelo de ayuda a los alumnos para escribir mejores revisiones literarias.
Petkov y Petkova (2006)	40 estudiantes de grado. Sin	Gestión de información económica	Distintas rúbricas en relación con	Un grupo trabajó sin rúbrica y el otro con ella. Al acabar el curso entregaron el proyecto final.	Se planteó la hipótesis nula "La media de las calificaciones del grupo control será más alta que la del grupo tratamiento", siendo rechazada.

REVISIÓN SOBRE AUTOEVALUACIÓN EDUCATIVA: EVIDENCIA EMPÍRICA DE SU IMPLEMENTACIÓN A TRAVÉS DE LA AUTOCALIFICACIÓN SIN CRITERIOS DE EVALUACIÓN, RÚBRICAS Y GUIONES

	información sobre edad.		la tarea.		Los autores concluyeron que la rúbrica mejoró el aprendizaje.
Sadler y Good (2006)	4 grupos de ciencias. 95 participantes. Sin datos para edad.	Biología: unidad didáctica sobre la clasificación de los organismos.	Rúbrica creada con los alumnos. Pero no se especifica si también el grupo control participó.	Hicieron un examen de Biología y dos de los cuatro grupos evaluaron a los pares y uno de los grupos se autoevaluó. Una semana después se repitió el mismo examen. En todos los grupos el profesor era el mismo.	Correlación alta entre las notas del profesor y las de la autoevaluación, y también fue alta con las notas de la evaluación de los pares pero más baja. Cuando se autoevaluaban alumnos con notas bajas tendieron a inflar su nota. (Esto para el primer examen). En el segundo examen, sólo el grupo que autoevaluó mejoró significativamente. La conclusión, es que tanto la autoevaluación como la evaluación de pares se puede usar para ahorrar tiempo al profesor, pero si se busca una mejora en el aprendizaje hay que usar la autoevaluación.
Schamber y Mahoney (2006)	Sin datos.	Curso general: "Today's decisions"	5 rúbricas para cada parte del trabajo a entregar.	Se trabajó en grupo para escribir el trabajo con cinco apartados. En el 2002 los estudiantes hicieron borradores en dos partes y en 2003 en las 5 partes.	Hacer borradores y la rúbrica fueron buenas estrategias para fomentar en los alumnos el pensamiento crítico.
McCormick et al. (2007)	44 universitarios en su último año. Sin datos de edad.	Programa de desarrollo del liderazgo.	Rúbrica	La rúbrica se usó para autoevaluarse, siendo la autocalificación comparada con la nota final del curso.	Los estudiantes demostraron mayor conocimiento del contenido, no basado en la rúbrica, pero en el mejor marco teórico del curso. La mitad de los alumnos fueron exactos en su autocalificación, mientras que la otra mitad se calificaron bajos.
Andrade, Du y Wang (2008)	116 estudiantes 46 de 3° y 70 de 4° grado. Edad no especificada.	Redacción y escritura	Rúbrica para autoevaluar la redacción	Los dos grupos tenían que escribir una redacción. El grupo experimental antes de hacerlo leyó una redacción modelo. Discutieron sus defectos y virtudes y crearon una lista con las cualidades de una buena redacción. Además recibieron una rúbrica para autoevaluar sus primeros borradores.	No hubo interacción entre la calidad de la redacción y el tiempo empleado. Tampoco hubo diferencias atendiendo al uso previo de rúbricas (los grupos estaban igualados). El experimental obtuvo mejores resultados en la redacción final.
Andrade et al., (2009)	Trabajo descriptivo sin grupo control.				
Andrade, Wang, Du y Akawi (2009)	268 estudiantes de Secundaria. Sin datos sobre edad.	Redacción y escritura	Rúbrica para autoevaluar la redacción	Los dos grupos tuvieron que realizar una redacción. Se trabajó durante cuatro sesiones. El grupo experimental antes de hacerlo leyó una redacción modelo. Discutieron sus defectos y virtudes y crearon una lista con las cualidades de una buena redacción. Además recibieron una rúbrica para autoevaluar sus primeros borradores. A los dos grupos se les pasó un cuestionario de autoeficacia sobre redacción. Se hizo en tres momentos: al inicio, en la segunda sesión después de que la rúbrica fuese entregada y en la cuarta tras la autoevaluación.	¿Hay efecto del uso a corto plazo de las rúbricas en la autoeficacia? Sólo apoyo parcial. Los participantes mejoraron su autoeficacia pero fue más un efecto de la práctica que también se dio en el control. ¿Hay efecto del uso a largo plazo de las rúbricas en la autoeficacia? No hubo efecto significativo del uso previo de rúbricas para los chicos pero si para las chicas. ¿Hay diferencia de género en autoeficacia? Las chicas tuvieron mejores resultados. ¿El efecto del tratamiento interactuó con el género? Sí, pues las chicas del grupo experimental obtuvieron en la tercera evaluación mejores resultados. No hubo esa interacción en los chicos. Una explicación para las diferencias de género es que las chicas tienen más satisfacción y confianza en su evidencia –autogenerada– de progreso al redactar que los chicos que necesitan evaluación externa. Además, las chicas tienden más a atribuir sus fracasos a aspectos de habilidad, mientras que los chicos a motivación y esfuerzo. Por lo tanto, generar la retroalimentación por ellas mismas ayuda a las chicas para protegerse de los efectos

					negativos de la retroalimentación de los adultos y/o compañeros.
Reitmeier y Vrchota (2009)	35 universitarios último curso. Rúbrica N=17. Hoja de reflexión N=18. Sin datos de edad.	Presentación oral grupo de Ciencias de Alimentación.	Rúbrica para la presentación oral	Se dividieron en dos grupos, unos usando rúbrica y los otros la hoja de reflexión para autoevaluar su presentación. Los profesores y los pares también evaluaron la presentación.	No se encontraron diferencias significativas entre los grupos, así que no hubo diferencia entre usar la rúbrica y la hoja de reflexión. Tampoco hubo diferencias basadas en quién evaluaba (el par, el propio alumno o el profesor) aunque las calificaciones de los pares fueron las más altas. Las calificaciones derivadas de la autoevaluación y las calificaciones de los profesores fueron muy parecidas. A partir del análisis de los comentarios de los alumnos, los autores concluyeron que la rúbrica es buena para exponer los conceptos pero que para los alumnos que alcanzan el nivel deseado usar la hoja de reflexión les permitía hacer reflexiones más profundas sobre la actividad.
Andrade, Du y Mycek (2010)	162 estudiantes. 75% chicas. Sin datos sobre edad.	9 grupos de Inglés	Rúbrica para redactar	La tarea era escribir una redacción. Los participantes en el grupo de comparación crearon una lista de criterios, escribieron un primer borrador, recibieron retroalimentación del profesor y entregaron la versión final. Los participantes del grupo de tratamiento además vieron un ejemplo, se les dio una rúbrica y la usaron para autoevaluar.	El grupo tratamiento obtuvo mejores calificaciones. El efecto del género no fue significativo.
Reynolds-Keefer (2010)	45 universitarios	Escritura	Rúbrica para escribir	Los participantes usaron la rúbrica para hacer dos trabajos de escritura. La rúbrica fue discutida en grupo dos semanas antes de la entrega de los trabajos. Los trabajos calificados fueron devueltos con una rúbrica que completa el profesor. Después los alumnos completaron un cuestionario de 9 ítems sobre el uso de las rúbricas.	Los estudiantes consideraron la rúbrica como útil, aunque algunos de ellos lo usaron como una lista de elementos que realizar. Para fomentar que esto no pase la rúbrica debe ser usada en el marco de evaluación formativa. Los estudiantes entendieron mejor las notas y se sintieron más justamente evaluados y que se trabajo fue mejor. 13 de los 45 alumnos manifestaron que su ansiedad bajó (pero no se informa del resto de estudiantes). El 60% manifestaron que utilizarían rúbricas cuando se convirtiesen en profesores. El 40% que dijo no lo basó en que (a) no sabían lo suficiente sobre rúbricas para utilizarlas en sus clases y (b) porque las rúbricas no son adecuadas para todos los niveles ni asignaturas.

Tabla 2. Estudios sobre los efectos de rúbricas

REVISIÓN SOBRE AUTOEVALUACIÓN EDUCATIVA: EVIDENCIA EMPÍRICA DE SU IMPLEMENTACIÓN A TRAVÉS DE LA AUTOCALIFICACIÓN SIN CRITERIOS DE EVALUACIÓN, RÚBRICAS Y GUIONES

Estudio	Diseño	Variables dependientes	Efecto de la intervención	Problemas (diseño, validez o datos)
Goodrich Andrade (2001)	Cuasi experimental con control	Calificaciones	Sí	No
Schafer et al. (2001)	Experimental con control	Aprendizaje de los alumnos a través del entrenamiento de sus profesores	Sólo para algunas medidas	Sí
Andrade y Boulay (2003)	Cuasi experimental	Calificaciones	No	No
Hafner y Hafner (2003)	Cuasi experimental sin control	Calificaciones	No se trata	Sí
Brown et al. (2004)	Cuasi experimental sin control	Calificaciones	-	Sí
Andrade y Du (2005)	Descriptivo	Percepción de los alumnos sobre rúbricas	-	No
Green y Bowser (2006)	Descriptivo	Realización de una revisión literaria	-	Sí
Petkov y Petkova (2006)	Cuasi experimental con control	Calificaciones	Los autores mantienen que sí pero faltan datos	Sí
Sadler y Good (2006)	Cuasi experimental con control	Calificaciones y calibración de las mismas	Sí	No
Schamber y Mahoney (2006)	Cuasi experimental sin control	Calificaciones	En duda	Sí
McCormick et al. (2007)	Cuasi experimental sin control	Aprendizaje y calibración de la autocalificación	Los autores mantienen que sí pero faltan datos	Sí
Andrade et al. (2008)	Cuasi experimental con control	Tiempo empleado para hacer la redacción, calificación	El tiempo no tuvo efecto, el uso de la rúbrica sí	Sí, porque hubo más intervención a parte de la rúbrica (exemplar)
Andrade, Buff et al. (2009)	Descriptivo	-	-	-
Andrade, Wang et al. (2009)	Cuasi experimental con control	Auto-eficacia	Sí	No
Reitmeier y Vrchota (2009)	Cuasi experimental sin control	Calificaciones	Sí	Sí

Andrade et al. (2010)	Cuasi experimental con control	Calificaciones	Sí	Sí, porque hubo más intervención a parte de la rúbrica (exemplar)
Reynolds-Keefer (2010)	Descriptivo	Percepción de los alumnos (replicación de Andrade y Du, 2005)	-	No, pero hay pequeños problemas en la presentación de los datos

Tabla 3. Análisis-resumen de los estudios sobre los efectos del uso de rúbricas

Primera, su uso no resulta sencillo para los alumnos (Andrade, Wang, et al., 2009; Goodrich Andrade, 2001). De esta forma, si los alumnos las utilizan en un primer momento es pensando que contienen lo que el profesor quiere que se haga (Andrade y Du, 2005). Esta sensación se ve reducida si, como se recomienda, los alumnos construyen las rúbricas con el profesor. Segunda, parece haber diferencias debidas al género que pueden suponer problemas para su implementación. Las alumnas las reciben peor al incluir comparación social (Andrade et al., 2008), si bien hay un estudio que encontró que aumentaba su sensación de autoeficacia (Andrade et al., 2009). Derivado de este hecho, creemos que es posible que las rúbricas orienten al alumno al resultado como ocurre con la autocalificación y, que su utilización deba ser seguida por preguntas más reflexivas como recomienda la investigación previa (Reitmeier y Vrchota, 2009). El hecho de que orienten al resultado sería debido a su carácter calificador y al hecho de que no aportan información sobre el proceso, sino sólo una evaluación centrada en el resultado final, aunque eso sí, a diferencia de la autocalificación, basada en criterios. Tercera, la rúbrica no garantiza con su entrega y uso el aprendizaje (Goodrich Andrade, 2001; Goodrich Andrade y Boulay, 2003; Jonsson y Svingby, 2007; Panadero y Jonsson, 2013), para que “funcionen” es necesario que se den las condiciones recomendadas (Goodrich, 1996) y se fomenten las actividades metacognitivas (Panadero y Jonsson, 2013). Y, cuarta, la gran mayoría de los estudios realizados con rúbricas presentan debilidades en sus diseños o tienen

afectada su validez. Como se puede comprobar en la Tabla 2, muchos de estos estudios no presentan suficientes datos estadísticos, siendo llamativo el alto número de estudios que no informan sobre la muestra. Además, un gran número de ellos son intervenciones sin medida de control, por lo que la ganancia debida a la intervención no se puede mantener, error que comenten varios de estos trabajos. Y, por último, la gran mayoría son diseños cuasi experimentales pues se trabaja con grupos naturales de aula que no permiten la asignación aleatoria de los participantes a las condiciones. Por ello, los resultados de estos estudios tienen que ser tomados con precaución.

En conclusión, las rúbricas tienen efectos positivos en la autoevaluación y el aprendizaje, pero sus efectos sobre la autorregulación no han sido investigados presentando, además, los problemas que se acaban de exponer. También hay dos efectos que no se han estudiado, ¿orientan las rúbricas al resultado al estar centradas en el producto? Y, ¿son las rúbricas los mejores instrumentos para fomentar la autorregulación? Para comprobar si habría algún tipo de ayuda que orientase a evaluar este proceso se analizaron los efectos de los instrumentos que se explican en la siguiente sección.

3.3. Autoevaluación a través del uso de guiones de autoevaluación (“scripts”, “prompts” y “cues”)

Los guiones de autoevaluación son ayudas que se proporcionan durante el proceso de una

actividad, normalmente en forma de pregunta, para sugerir al alumno que la responda para sí. Se diferencian de las pistas o indicios (*prompts* y *cues*) porque éstas son preguntas más o menos aisladas sobre aspectos específicos del proceso,

normalmente de tipo metacognitivo, mientras que los guiones presentan las preguntas de forma estructurada en categorías siguiendo el modelo experto de ejecución de la tarea (Tabla 4).

**GUIÓN PARA AUTOEVALUAR
EL PROCESO DE CONSTRUCCIÓN DE MAPAS CONCEPTUALES**

1. ¿Me he hecho una pregunta de enfoque? (¿Tengo claro qué es lo que debe recoger el mapa conceptual?)
2. Basándome en esa pregunta, ¿he identificado conceptos suficientes para recoger las ideas del material trabajado?
3. ¿He ordenado los conceptos jerárquicamente poniendo los más amplios y globales al principio de la lista? (OJO: dependiendo de la naturaleza del mapa, al comienzo pueden ir conceptos concretos y específicos como, por ejemplo, si el mapa va a hablar del agua)
4. ¿He revisado la lista para añadir más conceptos si era necesario?
5. ¿He comenzado a construir el mapa con los conceptos más globales (o los conceptos clave que hay que explicar) en la parte superior?
6. ¿He colocado 3 ó 4 sub-conceptos debajo de los más globales? Ten en cuenta que si tienes más de 6 conceptos en un mismo nivel es posible poner un concepto intermedio creando un nuevo nivel jerárquico (aunque no siempre es posible pues, por ejemplo, hay 8 clases de insectos).
7. ¿He unido los conceptos mediante líneas utilizando nexos –palabras que se ponen en las líneas para dar más información de la relación–?
8. ¿He dibujado líneas que relacionasen (vinculasen) conceptos situados en diferentes columnas de la jerarquía? Ayudan a descubrir nuevas relaciones.
9. ¿He puesto ejemplos bajo los conceptos para que fueran más sencillos de visualizar?
10. ¿He repasado el mapa conceptual una vez hecho teniendo en cuenta...
... que hay muchas formas de organizar los conceptos?
... cuál era el propósito perseguido y la información con que contaba?

Tabla 4. Ejemplo de guión de autoevaluación para la construcción de mapas conceptuales. Autores: Ernesto Panadero, Jesús Alonso-Tapia y Juan Antonio Huertas

Dado que estos últimos siguen el modelo experto, resultan más específicos que las pistas o indicios pues hacen referencia a aspectos de la tarea que se va a realizar. En cualquier caso, el objetivo de todas ellas es promover la reflexión sobre la adecuación del proceso que se va realizando y sobre los resultados parciales que

se van obteniendo. Estos instrumentos también permiten autoevaluar el trabajo al acabar pero, si sólo se usan en ese momento, pierden parte de su potencial pues los alumnos tendrían que hacer de nuevo la actividad en caso de detectar que algo es incorrecto (Bannert, 2009; Thillmann, Kunsting, Wirth, y Leutner, 2009).

Como ocurría con las rúbricas, también en relación con los guiones de autoevaluación, pistas e indicios cabe preguntarse: ¿son adecuados para facilitar la autoevaluación, la autorregulación y el aprendizaje?

Se ha encontrado que las pistas, los indicios y los guiones tienen muchos aspectos positivos, como muestran los trabajos presentados en la Tabla 5.

Estudio	Muestra	Tarea	Tipo de ayuda	Procedimiento	Resultados e interpretaciones
Dreher y Brown (1993)	34 universitarios recompensados con 10 \$. Edad no especificada.	Búsqueda de términos en un libro de texto	Pista: tres frases que recuerdan de forma general la planificación	Los participantes entraban de forma individual. Hacían un test previo de conocimientos en Psicología -pues el libro de texto era de Psicología- y se evaluaba su rapidez y exactitud de lectura. La actividad consistía en realizar una búsqueda de términos en el libro con un máximo de cinco minutos. Los que estaban en la condición con pista se les leía ésta al principio de cada búsqueda. Hay que tener en cuenta que los términos para buscar podían estar indexados o no (es decir presentes en el índice). Los autores consideran que estar o no indexado es una variable intra-sujeto, mientras que estar en el grupo con pista o no sería una variable inter-sujeto. Las dependientes serían el número de respuestas correctas al término por el que se les preguntaba y el tiempo de búsqueda.	El uso de la pista antes de cada búsqueda activó la planificación de la misma, especialmente en los casos donde el término estaba indexado. En los casos en los que el término no estaba indexado, el uso de la pista aumentó la búsqueda en páginas clave mientras que redujo la cantidad total de páginas buscadas.
Elischberger y Roebbers (2001)	83 participantes. 37 de Educación Infantil edad 71,6 meses. 46 de 2º grado edad 97,6 meses	Recordar un video en el que se roba a un niño	Pistas verbales	Los participantes veían el vídeo en grupos de 5 ó 6. Tres semanas después se les entrevistaba de forma individual. Se les pedía que dijese todo lo que recordasen del vídeo. Cada vez que el alumno permanecía callado o decía que no se acordaba de más se daba por terminada esa parte. Al grupo sin pistas se les pedía tres veces que dijese todo lo que recordasen sobre el video. Al grupo con pistas se les dijo primero que dijese todo lo que recordaban, después que dijese todo lo que recordaban haber oído y por último todo lo que recordaban haber visto. En eso consistían las pistas. Finalmente se les pasó un test de inteligencia verbal para estimar si los grupos estaban igualados.	Primero comprobaron si el nivel de recuerdo era igual tras la primera vez que se preguntaba. No hubo diferencias significativas por lo que los dos grupos estaban igualados en recuerdo antes de dar las pistas. Después se realizó una ANCOVA que permite mantener que el grupo con pistas recordó y reportó significativamente más aspectos del video que el grupo control. Por lo tanto las pistas verbales (en este caso son muy sencillas) ayudan a recordar. Además la exactitud del recuerdo no empeoraba. Un dato significativo: en el grupo que recibió pistas 11 de 17 niños de la guardería y 18 de 25 de 2º grado dijeron que añadieron recuerdos la segunda vez que se les preguntó frente a 3 de 17 y 8 de 21 en el control. Los resultados sobre inteligencia verbal muestran que tuvo un efecto significativo sobre la cantidad de recuerdos pronunciados (recordar que ambos grupos estaban igualados en inteligencia verbal).
Ge y Land (2001)	115 estudiantes de Grado Edad no especificada.	Curso introductorio Ciencias de la Información y Tecnología	Entrega de pistas y/o trabajo cooperativo	Se asignaron las condiciones de acuerdo a los grupos en los que se dividía la clase, supuestamente de forma aleatoria. Primero atendían a una clase de 75 minutos que daba el catedrático y después trabajan durante 115 minutos con los profesores	El grupo individual con pistas demostró mejores capacidades para resolver problemas que aquellos que no recibían pistas. Se hipotetizó que los alumnos trabajando en grupo, con o sin pistas, demostrarían mejores habilidades para la resolución de

REVISIÓN SOBRE AUTOEVALUACIÓN EDUCATIVA: EVIDENCIA EMPÍRICA DE SU IMPLEMENTACIÓN A TRAVÉS DE LA AUTOCALIFICACIÓN SIN CRITERIOS DE EVALUACIÓN, RÚBRICAS Y GUIONES

				ayudantes todos iguales en experiencia.	problemas. Hubo diferencias en el grupo cooperativo con pistas sobre los grupos individuales, pero no las hubo en el grupo cooperativo sin pistas sobre el grupo individual sin pistas. Los alumnos trabajando en grupo y con pistas demostraron las mejores habilidades de todos.
Kramarski y Zeichner (2001)	186 estudiantes de 11 Grado. Media de edad: 17,5. 8 clases naturales.	Matemáticas .	El grupo MF recibió al final de la tarea pistas sobre razonamiento matemático y RF pistas sobre el resultado	Se evaluó a los estudiantes con un test de 27 ítems, después resolviendo los problemas trabajaron con el ordenador que les daba la retroalimentación. Después repitieron el test del inicio.	Los participantes con MF obtuvieron mejores resultados a parte de otras medidas relacionadas con su aprendizaje.
Atkinson, Renkl y Merrill (2003)	Exp. 1 78 estudiantes de Psicología. Edad no especificada. Exp. 2 40 estudiantes de Secundaria. Edad no especificada.	Exp. 1 Problemas de probabilidad Exp. 2 Problemas de probabilidad	Exp. 1 La pista consistía en animar al participante a autoexplicarse el ejercicio que acababa de hacer. Exp.2 Igual que en el exp. 1.	Exp. 1. Los alumnos realizaban ejercicios de probabilidad con un ordenador. Los grupos que tenían pista además la recibían. Exp. 2 Idéntico al Exp. 1.	Exp. 1 El grupo con ejercicios hacia atrás que desaparecían (en blanco) obtuvo mejores resultados aumentando además la transferencia cercana y la lejana. Las pistas tuvieron un efecto significativo mejorando el aprendizaje, y no aumentaron significativamente el tiempo empleado. No hubo interacción significativa entre ejercicios que desaparecían y pistas. Exp. 2 La intervención fue significativa para el aprendizaje y además para que los alumnos hiciesen mejor los ejercicios que eran estructuralmente diferentes de los que habían recibido instrucciones. Por lo tanto, el uso de pistas fomentó el aprendizaje y la transferencia a ejercicios diferentes.
Winne y Jamieson-Noel (2003)	69 estudiantes de Grado. Edad 21.73. 18 chicos 51 chicas. Recibieron 15 \$ y entraron en un sorteo de 4 premios de 50\$.	Capítulo sobre meteorología	Había pistas embebidas en el texto (por ejemplo enlaces a páginas web).	Los participantes se familiarizaron con el programa, después durante 20-30 minutos estudiaron el capítulo pudiendo visualizar 4 objetivos de instrucción: (a) describir conceptos importantes, (b) explicar las relaciones de causa efecto, (c) aplicar los principios del fenómeno, (d) buscar soluciones alternativas. Además había 5 tipos de pistas embebidas. Cuando acabaron de estudiar hicieron un cuestionario sobre la utilización de 7 estrategias y si las habían usado para alcanzar los objetivos, así como una prueba de rendimiento.	Los autores compararon la frecuencia reportada por los participantes a la estimación de su modelo teórico, siendo la de los participantes más baja así como la utilidad que le otorgaban a las estrategias. Concluyeron que los objetivos influyeron en lo que los participantes estudiaron pero no en cómo lo hicieron. En suma, los participantes no estudiaron como predecía el modelo lo que conlleva que puede ser difícil cambiar las perspectivas de los alumnos dado que ellos dotan de diferente utilidad a las estrategias con respecto al modelo propuesto.
Van den Boom et al. (2004)	42 estudiantes 2º Magisterio. Edad: Media 20,1 Desviación típica: 1,5 Recibieron 30€	Psicología de la Salud	Pistas metacognitivas vs. No metacognitivas. Retroalimentación del tutor vs. No retroalimentación	Se les pasó un cuestionario de autorregulación. Después trabajaban con el ordenador en 3 actividades. Tras la primera recibían la pista y retroalimentación que les correspondía. El grupo que no tenía pasaba directamente a la 2ª. Al acabar realizaron el cuestionario de autorregulación y se evaluó el aprendizaje.	La retroalimentación aumentó significativamente la autorregulación. Las pistas no mejoraron la autorregulación aunque sí en interacción con la retroalimentación. El aprendizaje no presentó efecto. Hipotetizan que debido al poco tiempo de intervención. La opinión de los participantes sobre las pistas y retroalimentación del tutor es positiva. Los participantes cuyas pistas no fueron metacognitivas y no tuvieron retroalimentación evaluaron sus pistas como disturbadoras.
Ertl,	Estudio 1: 24	Estudio 1:	El apoyo	Estudio 1: los participantes del grupo	Estudio 1: no hubo efecto del apoyo

<p>Fischer y Mandl (2006)</p>	<p>Estudio 2: 86 Estudio 3: 159</p>	<p>resolución de problemas sobre motivación Estudio 2: enseñanza cooperativa sobre efectos ambientales del genotipo Estudio 3: resolución de problemas sobre la teoría de la atribución</p>	<p>conceptual es un esquema con conceptos teóricos, evidencia, consecuencias y opinión personal Las prompts sirven para ayudar en el trabajo entre las parejas pues lo estructuran</p>	<p>con visualización de ayudas estructurales (esquemas que rellenar) Estudio 2: Uno de los miembros de la pareja hace de profesor y el otro de alumno. Comparten las conclusiones. Estudio 3: parecido al estudio 2 sólo cambiando la tarea.</p>	<p>conceptual. Estudio 2: el <i>script</i> centra al alumno en los conceptos teóricos, mientras que el apoyo conceptual centra en las elaboraciones personales negando los conceptos teóricos, pero no tuvo efectos en el aprendizaje. Estudio 3: el apoyo conceptual influyó en el entorno colaborativo y la adquisición de conocimiento a nivel individual. El apoyo sociocognitivo (<i>prompt</i>) mostró un efecto pequeño. La conclusión es que las <i>prompts</i> es mucho más efectiva en combinación con el apoyo conceptual.</p>
<p>Berthold, Nückles y Renkl (2007)</p>	<p>84 estudiantes de 1º Psicología Edad no especificada.</p>	<p>Cambio conceptual, Psicología Evolutiva</p>	<p>Entrega de pistas cognitivas, metacognitivas, mixtas o ausencia.</p>	<p>Hubo cuatro condiciones. A las tres que tenían pistas se les entregó para el trabajo. Asisten durante dos sesiones a ver un video sobre cambio conceptual. Después crean un protocolo de aprendizaje tras recibir las instrucciones que corresponden a su condición. Al acabar contestaron un cuestionario sobre la ayuda percibida, un post-test para evaluar su aprendizaje y se autoevaluaron.</p>	<p>a) Estrategias cognitivas y metacognitivas: a.1. Las condiciones con pistas cognitivas y mixtas presentaron diferencias significativas con el grupo control en estrategias cognitivas. No en la condición metacognitiva. a.2. Todas las condiciones con pistas presentaron diferencias significativas en estrategias metacognitivas. b) Aprendizaje: pistas cognitivas y mixtas presentaron diferencias significativas con el control. No en la condición metacognitiva. c) Precisión de la autoevaluación sin efecto. d) Ayuda percibida: no hubo efecto.</p>
<p>Grofle y Renkl (2007)</p>	<p>Experimento 1: 118 Estudiantes. 84 mujeres y 34 hombres. Edad media: 23,2.</p>	<p>Exp. 1: Probabilidad</p>	<p>Pistas metacognitivas sobre el proceso (3 preguntas). Pistas diferentes para los tres grupos. El momento de presentación de las pistas no se especificó.</p>	<p>Experimento 1: se hizo una medida pre con problemas de probabilidad en la que no se encontraron diferencias. También se controló por ejecución previa en Matemáticas. Después, 8 ejemplos se presentaron siendo estos diferentes para los grupos. Después hubo un post-test con 15 problemas (4 de transferencia cercana y 11 lejana). Experimento 2: no es relevante para los objetivos de la revisión.</p>	<p>Experimento 1: No hubo efecto de las pistas. Los autores lo explicaron porque era una actividad compleja y las pistas eran "solo" tres preguntas metacognitivas.</p>
<p>Kollar, Fischer y Slotta (2007)</p>	<p>90 estudiantes de Secundaria, 8º-10º grado. Edad: 15,3.</p>	<p>"El misterio de la rana deforme". Conocimiento del Medio</p>	<p>Entrega de guión externo alto o bajo en estructuración. El guión se centraba en cómo discutir y argumentar los aspectos importantes de la actividad.</p>	<p>Primero se evaluaron los guiones internos de los alumnos a través de cuestionarios durante dos sesiones. Podían ser guiones internos con nivel alto o bajo de estructuración. Dos semanas después se crearon parejas homogéneas atendiendo al nivel de guión interno. Trabajaron en la unidad en el ordenador recibiendo un guión externo con un nivel de estructuración bajo o alto dependiendo del grupo. Con un test sobre argumentación se evaluó el dominio general con un máximo de 12 puntos y con un test con cinco preguntas abiertas para el dominio específico.</p>	<p>a) Los alumnos con niveles internos más altos de guión crearon más nuevos argumentos que los que los tenían bajos. b) Las parejas con niveles internos altos también produjeron más argumentos, contraargumentos, y argumentos integradores. c) Con respecto a la adquisición de conocimiento general de dominio fueron de más a menos: guión interno alto-externo alto // interno bajo-externo alto // interno algo-externo bajo // interno bajo-externo bajo. d) Adquisición de conocimiento específico de dominio: los participantes con niveles internos más altos consiguieron mejores resultados, especialmente en interacción con el guión externo alto. En conclusión, los guiones externos aumentaron el aprendizaje en tareas</p>

REVISIÓN SOBRE AUTOEVALUACIÓN EDUCATIVA: EVIDENCIA EMPÍRICA DE SU IMPLEMENTACIÓN A TRAVÉS DE LA AUTOCALIFICACIÓN SIN CRITERIOS DE EVALUACIÓN, RÚBRICAS Y GUIONES

					colaborativas, si bien es cierto que los que más se beneficiaron fueron los alumnos con guiones internos altos. Esto contradice la hipótesis de que individuos con niveles internos altos podrían no sacar provecho de los guiones externos altos pues entrarían en conflicto con su guión interno (es decir, que quién tiene ya estructurado su guión de trabajo interno tendría interferencias al seguir un guión externo alto). Resultó ser al contrario: los que más se favorecieron del guión externo alto fueron los que tenían el interno también alto.
Schworn y Renkl (2007)	72 estudiantes de Magisterio. Media de edad 23,4. Recibieron 15 €	Dos videos sobre la investigación de células y las diferencias por género en Matemáticas y Ciencias Naturales	Tres tipos: a) 8 pistas basadas en ejemplos auto-explicativos b) 8 pistas basadas en aprendizaje autoexplicativo c) una mezcla de las dos anteriores	Los alumnos trabajaron en sesiones de tres horas. Se les hicieron varios pretest para comprobar si los grupos eran iguales en: mecanografía, conocimiento sobre el contenido de los videos y habilidades de argumentativas. Después veían los videos. Primero enteros sin interrupciones, después deteniéndose en 4 momentos diferentes para que los alumnos argumentaran y escribieran sus respuestas. En estas interrupciones aparecían las pistas en las condiciones que las tenían. <u>En estas condiciones no se podía continuar hacia la nueva sección a menos que hubiesen escrito una auto-explicación mientras que en la condición control si se podía pasar.</u> Luego completaban 4 cuestionarios: (a) sobre el contenido del video, (b) sobre conocimiento declarativo de argumentación, (c) sobre habilidades argumentativas, (d) aceptación de la intervención.	Conclusiones: a) Es viable usar ejemplos audiovisuales de doble contenido para potenciar el aprendizaje de conocimiento declarativo sobre argumentación. b) Las pistas auto-explicativas favorecen las habilidades argumentativas y mejoran la calidad de las auto-explicaciones. c) Las pistas referidas al dominio del aprendizaje potencian el número y calidad de las declaraciones de referidas al dominio del aprendizaje. Sin embargo, no hubo resultados significativos en el aprendizaje del dominio. e) Las pistas referidas al aprendizaje no fueron percibidas por los participantes como positivas aunque potenciaron el aprendizaje de habilidades argumentativas. Sin embargo, los participantes no fueron informados de las ventajas de las pistas.
Kostons, van Gog y Paas (2009)	40 trabajadores de la Universidad. 26 mujeres, 14 hombres. Edad: 36,62, SD=10,78.	Leyes de Mendel	La pista consistía en una muestra del movimiento de los ojos de los participantes en la pantalla. Se mostró al final de la tarea, en la fase de auto-reflexión.	Los participantes realizaron un test sobre su conocimiento de las leyes de Mendel para asignarles a un grupo (alta o baja experiencia.), y después fueron asignados de forma aleatoria a un grupo con o sin pistas. Hicieron la tarea en un ordenador sobre las leyes de Mendel. El grupo sin pistas se autoevaluó en voz alta, mientras que el grupo con pistas también autoevaluó pero primero visualizaron el video con su movimiento de ojos.	La pista fue útil para los “novatos” pero no para los “expertos. Esto apoya la Teoría de la Carga Cognitiva dado que el sistema cognitivo de los novatos estaba “saturado” y no fueron conscientes de su ejecución. Sin embargo, los novatos no lo hicieron mejor en las categorías de monitorización/ evaluación, posiblemente porque les faltaban los criterios para hacerlo de forma correcta. También se prueba la interacción entre autoevaluación y experiencia previa: los expertos tuvieron menos carga cognitiva y por ello fueron capaces de autoevaluar: usaron más criterios, hicieron más afirmaciones positivas (pero no se comprobó la exactitud de las mismas).
Kramarski y Dudai (2009)	100 estudiantes de 9º grado. 47 chicos, 53 chicas. 3 grupos naturales.	Problemas matemáticos	Dos tipos de pistas: SEG preguntas sobre la propia ejecución y CFG pregunta sobre la ejecución de los compañeros. Las pistas se mostraron durante	Los participantes hicieron un test de Matemáticas que no mostró diferencias. Después los grupos experimentales recibieron formación sobre cómo responder a las pistas. Después los participantes ejecutaron la tarea en cuatro sesiones, trabajando en grupos de cuatro. La evaluación tras la intervención se realizó a través de una actividad de transferencia	“Aprendizaje matemático”: los alumnos CFG superaron a los SEG mientras que estos superaron a los alumnos control. “Retroalimentación matemática”: 5 categorías en dos CFG superaron a los SEG, y los SEG superaron en otras dos a los CFG. “Transferencia de las habilidades matemáticas”: 4 categorías. En tres de ellas CFG superaron a SEG. En dos de

			la solución del proceso, por lo tanto en la fase de auto-reflexión, en un ordenador con ventanas emergentes.	cercana y otra lejana, un test escrito y el cuestionario de autorregulación.	esas tres SEG superó al control. "Autorregulación": 3 categorías. En dos de ellas CFG superaron a SEG y control. "Retroalimentación metacognitiva": 4 categorías. En tres de ellas CFG superaron a SEG y en la cuarta fue a la inversa.
Kramarski y Michalsky (2009)	144 profesores en su primer año. 62% mujeres, 38% hombres. Edad media: 22,6.	Diseño de unidades didácticas	3 tipos de pistas: (a) M_P recibió preguntas de comprensión en la fase de planificación; (b) M_A recibió preguntas estratégicas en la fase de ejecución; (c) M_E recibió preguntas reflexivas en la fase de auto-reflexión. Todo se recibió a través de ventanas emergentes en el ordenador.	La autorregulación y las habilidades de comprensión y diseño de unidades didácticas (TCPK) de los participantes fueron evaluadas antes de la intervención. Después, los participantes trabajaron durante 14 sesiones con las pistas correspondientes. Al final se evaluó autorregulación y TCPK de nuevo.	No hubo diferencias en la medida pre. La condición M_E superó a las otras dos en todas las medidas. M_P superó a M_A. Por lo tanto, parece que dar pistas en la fase final de la autorregulación fue la forma más efectiva de fomentar el aprendizaje y la autorregulación. El hecho de que M_A fuese la intervención más pobre se podría explicar desde la Teoría de la Carga Cognitiva: la fase de ejecución requiere altos niveles de atención y puede ser difícil procesar pistas en esa fase.
Alonso Tapia y Panadero (2010)	80 alumnos 3º ESO 50% chicas 50% chicos Rango de edad 15-16	Análisis de paisajes	Guión de autoevaluación con el proceso experto de realización de un análisis de paisaje	Se pasó el cuestionario de orientación de motivación sin encontrar diferencias. De forma individual analizaban 3 paisajes recibiendo las instrucciones, retroalimentación y el guión acordes con la condición completando un cuestionario de habilidad percibida. Analizaban diciendo en voz alta sus pensamientos, escribiendo un resumen del análisis y después se daba retroalimentación. Se repetía para los dos paisajes restantes. Al acabar completaban de nuevo el cuestionario de habilidad percibida y uno de autorregulación.	El uso del guión fomentó las conductas autorregulatorias positivas y disminuyó las negativas. En interacción con la retroalimentación al proceso disminuyó las conductas autorregulatorias negativas. Las instrucciones orientadas al aprendizaje también disminuyeron las conductas autorregulatorias negativas. Los alumnos con guión demostraron un aumento significativo de su aprendizaje frente a los que no tenían esta ayuda. El guión no tuvo un efecto significativo en la habilidad percibida, seguramente por la corta duración de la intervención, pero sí se encontró una tendencia a aumentar en las condiciones con guión.
Kramarski y Michalsky (2010)	95 profesores de Ciencias en su primer año. 57 mujeres y 38 hombres. Edad media 24,5.	Diseño de unidades didácticas	Indicios metacognitivos. 4 categorías: comprensión, conexión, estrategia y preguntas reflexivas	Evaluación de conocimientos previos el primer día. Después los participantes trabajaron en parejas durante 14 sesiones (56 horas), implementando la enseñanza y los métodos de aprendizaje. Los participantes en el grupo de intervención tuvieron los indicios durante las sesiones. El último día se hizo la evaluación post-intervención. IMPORTANTE: aunque trabajaron en parejas la evaluación fue individual.	TPCK habilidades: el grupo HYP+META obtuvo mejores resultados. Autorregulación: el grupo HYP+META mejores resultados. También mejores resultados para la medida de reflexiones escritas. Conclusión: los indicios metacognitivos mejoraron la autorregulación y el aprendizaje.

Tabla 5. Estudios sobre los efectos de guiones, pistas e indicios

Entre otros: favorecen el aprendizaje (Alonso-Tapia y Panadero, 2010; Atkinson, Renkl, y Merrill, 2003; Berthold, Nückles, y Renkl, 2007; Dreher y Brown, 1993; Elischberger y Roebers, 2001; Ge y Land, 2001; Kollar, Fischer, y Slotta, 2007; Kostons, Van

Gog, y Paas, 2009; Kramarski y Zeichner, 2001; Schworn y Renkl, 2007; Van den Boom, Paas, Merriënboer, y Van Gog, 2004); mejoran la autorregulación cuando se utilizan solos (Alonso-Tapia y Panadero, 2010; Berthold et al., 2007; Kramarski y Dudai, 2009; Kramarski y

Michalsky, 2009, 2010), así como cuando su uso se combina con retroalimentación (Van den Boom et al., 2004); aumentan la capacidad de búsqueda de términos en textos a través de una mejora en la concentración, localización de elementos clave y planificación (Dreher y Brown, 1993); son recomendables porque (a) son fáciles de implementar, (b) no alargan el tiempo de aprendizaje y (c) potencian la transferencia (Atkinson et al., 2003); incrementan los elementos que se recuerdan sin perder exactitud desde muy temprana edad (Elischberger y Roebers, 2001); en interacción con el trabajo en grupo aumentan significativamente el rendimiento (Ge y Land, 2001); aunque el nivel de estructuración interna –la estructura mental para realizar un procedimiento– del individuo sea alta y el nivel de estructuración del guión también lo sea, no interfieren e interactúan para aumentar el aprendizaje (Kollar et al., 2007); y cuando se dan pistas distraen menos si son autorregulatorias (Van den Boom et al., 2004). De esta forma, el uso de pistas, indicios y guiones tiene efectos muy positivos a nivel educativo, principalmente, en la autorregulación y el aprendizaje.

Además de estos aspectos positivos, otra ventaja del uso de guiones es que no parece tener efectos negativos, aunque los estudios sobre su efectividad no siempre han dado resultados significativos. Así, Van den Boom et al. (2004) encontraron que el uso de pistas no aumentó el aprendizaje, si bien la intervención fue breve; Ertl, Fischer y Mandl (2006) encontraron efectos significativos de su intervención sólo en algunas de las medidas; Grofle y Renkl (2007) no encontraron un efecto significativo de su intervención; finalmente, Berthold et al. (2007) encontraron que el uso de pistas no aumentó la precisión en la autocalificación, si bien estos autores no informaron si la igualdad en la precisión entre grupos se debió a que todos los grupos fueron

muy precisos. Otro aspecto que hay que tener en cuenta en la investigación de guiones y pistas, es que la mayoría de los trabajos, como se puede comprobar en la Tabla 6 son de carácter experimental, con diseños sólidos, buen uso y presentación de los análisis estadísticos, en clara contraposición con los trabajos sobre el efecto de las rúbricas que son más “laxos” o incompletos en esos aspectos.

De esta forma, la solidez de los estudios sobre el efecto de los guiones es superior a la de los estudios sobre rúbricas, si bien los primeros se han centrado en diseños experimentales y no han sido puestos en práctica en el aula en situaciones naturales como sí se ha hecho con las rúbricas.

Tal vez uno de los mayores problemas de la investigación sobre guiones sea que no se ha tenido en cuenta que la efectividad de éstos radica en que son ayudas a la autoevaluación de la calidad del proceso así como la de los resultados obtenidos –parciales o totales–. Por esta razón, no se ha investigado sistemáticamente sus efectos sobre la autoevaluación especialmente en contextos naturales.

Desde nuestro punto de vista, los guiones pueden ser ayudas muy útiles para facilitar la reflexión y toma de conciencia sobre los pasos claves del proceso y, de esta forma, para avanzar en la actividad. El hecho de que centren la atención del alumno en el proceso y no exclusivamente en el producto como las rúbricas, es un valor añadido. Esto sin olvidar que también se pueden utilizar para la autoevaluación final, si bien no contienen el gradiente para la autocalificación, y la investigación indica que no es el procedimiento adecuado (Bannert, 2009; Thillmann et al., 2009), a menos que se trate de pistas o indicios que estén dirigidos a reflexionar sobre el proceso a posteriori (Kramarski y Michalsky, 2009).

Estudio	Diseño	Variables dependientes	Efecto de la intervención	Problemas (diseño, validez o datos)
Dreher y Brown (1993)	Experimental con control	Aprendizaje y autorregulación	Sí	No
Elischberger y Roebers (2001)	Experimental con control	Aprendizaje	Sí	No
Ge y Land (2001)	Experimental con control	Aprendizaje de resolución de problemas	Sí	No
Kramarsky y Zeichner (2001)	Cuasi experimental sin control	Aprendizaje	Sí	No
Atkinson, Renkl y Merrill (2003)	Experimental	Aprendizaje	Sí	No
Winne y Jamieson-Noel (2003)	Descriptivo	Aprendizaje y uso de estrategias	- (comparado contra el modelo teórico)	No
Van den Boom et al. (2004)	Experimental con control	Autorregulación, aprendizaje, opinión sobre la ayuda	Sí	No
Ertl, Fischer y Mandl (2006)	Experimental con control	Aprendizaje	No	No
Berthold, Nückles y Renkl (2007)	Experimental con control	Aprendizaje, uso de estrategias, calibración autocalificación y ayuda percibida	Sí, menos para las dos últimas	No
Grofle y Renkl (2007)	Experimental 2x3	Aprendizaje	No	No
Kollar, Fischer y Slotta (2007)	Experimental 2x2	Aprendizaje	Sí	No
Schworn y Renkl (2007)	Experimental con control	Aprendizaje y percepción alumnos	Sí, para el aprendizaje	No
Kostons, van Gog y Paas (2009)	Experimental con control	Criterio de autoevaluación usado	Sí	No
Kramarski y Dudai (2009)	Cuasi-experimental con control	Autorregulación y transferencia	Sí	No
Kramarski y Michalsky (2009)	Cuasi-experimental sin control	Autorregulación y aprendizaje	Sí	No
Alonso Tapia y Panadero (2010)	Experimental 2x2x2	Autorregulación, aprendizaje y autoeficacia	Sí	No
Kramarski y Michalsky (2010)	Experimental con control	Autorregulación y comprensión	Sí	No

Tabla 6. Análisis-resumen de los estudios sobre los efectos del uso de guiones, indicios y pistas

4. CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

¿Qué se puede concluir de la revisión de los estudios empíricos comparando la autoevaluación sin criterios, el uso de rúbricas y el uso de guiones? Una primera conclusión que se puede extraer de esta revisión es que los estudios para fomentar la autoevaluación sin especificar los criterios de evaluación presentan problemas para demostrar una ganancia de aprendizaje en los alumnos (Tan, 2012). Esto se debe a la heterogeneidad de los estudios que se denominan como autoevaluación cuando lo que fomentan en realidad es la autocalificación. También se puede concluir que los estudios realizados con rúbricas y guiones ofrecen más garantías de eficacia para ayudar a los alumnos a autoevaluarse. Por este motivo la primera conclusión, respondiendo a la primera pregunta de esta investigación, es que es preferible implementar la autoevaluación en el aula usando instrumentos que contengan los criterios de evaluación (Andrade y Valtcheva, 2009; Reynolds-Keefer, 2010).

Con respecto a la segunda pregunta de investigación –cómo se han utilizado estas ayudas y cuál es la evidencia de su efectividad–, nos centraremos a continuación en las rúbricas y los guiones, dados los problemas hallados en la efectividad de la autoevaluación sin criterios de evaluación (Falchikov y Boud, 1989; Tan, 2012).

Por lo que a las rúbricas se refiere, la investigación que hemos revisado sobre sus efectos presenta un punto débil. Muchos de los estudios sobre rúbricas tienen problemas de validez por, por ejemplo, problemas en el diseño, falta de datos sobre la muestra y datos estadísticos, etc. No obstante, a tenor de los resultados en los estudios más sólidos se puede mantener que las rúbricas tienden a fomentar la autoevaluación y el rendimiento, pero su entrega no garantiza el éxito *per se*, sino que es necesaria su combinación con actividades metacognitivas (Panadero y Jonsson, 2013).

En cuanto a los guiones, la investigación revisada ha puesto de manifiesto que el uso de los mismos tiende a fomentar la autoevaluación, la autorregulación y el aprendizaje dado que centran la atención del alumno en el proceso que hay que seguir (Alonso-Tapia y Panadero, 2010; Bannert, 2009; Bannert, Hildebrand, y Mengelkamp, 2009; Dreher y Brown, 1993; Ertl et al., 2006; Kostons et al., 2009; Kramarski y Michalsky, 2009, 2010; Kramarski y Zeichner, 2001; Montague, 2007). Sin embargo, presentan tres puntos débiles. Primero, la mayoría de los estudios han sido experiencias de laboratorio, siendo pocos los realizados en contextos naturales, por lo que la efectividad de su aplicación en el aula sigue siendo una incógnita. Segundo, se han centrado en gran medida en tareas con ordenadores sin mediar la presencia de un profesor. Y tercero, los efectos positivos no se han encontrado siempre pues parece que dependen de la calidad y estructura del guión y de la duración de la intervención (Berthold et al., 2007; Kitsantas, Reiser, y Doster, 2004; Kollar et al., 2007). Por lo tanto, es necesario analizar, por una parte, los efectos del uso de guiones en contextos controlados –de laboratorio– usando ordenadores con un profesor y, por otra parte, cuáles son sus efectos en situaciones naturales de aula.

Atendiendo a todo lo expuesto, una primera futura línea de trabajo sería investigar *sistemáticamente* –en distintas condiciones– la eficacia que los guiones puedan tener para facilitar la autoevaluación, tanto de la adecuación del proceso que hay que seguir, como la de los resultados que se van obteniendo, y sus efectos sobre la autorregulación, el aprendizaje y la propia motivación. En segundo lugar, analizar la eficacia sobre las mismas variables de las rúbricas, especialmente teniendo en cuenta que el efecto del uso de las rúbricas en la autorregulación no ha sido estudiado.

Un tercer aspecto para investigar tiene que ver con las diferencias entre rúbricas y guiones, diferencias que pueden dar lugar a que sus efectos sean distintos. Se diferencian fundamentalmente en dos aspectos. Primero, en el tipo de información que proporcionan al

alumno: las rúbricas contienen unas categorías muy específicas sobre el producto final que el alumno tiene que alcanzar, mientras que los guiones son preguntas que el alumno tiene que responder, preferentemente durante el proceso de trabajo. Esta diferencia podría dar lugar a que los alumnos se centrasen en alcanzar el producto final cuando usan rúbricas, y en el proceso que hay que seguir cuando usan guiones. Esto es, los alumnos se orientarían a metas diferentes dependiendo del instrumento que usen con el consiguiente efecto en el modo de trabajar. Segundo, se diferencian en la importancia que dan a la calificación. Las rúbricas contienen un baremo para que el alumno autocalifique su trabajo, lo que probablemente centra su atención en la nota, mientras que los guiones no contienen nada semejante, por lo que, posiblemente, centren la atención en el aprendizaje de la tarea que se esté realizando. Estas diferencias plantean dos preguntas. Primera, ¿tiene más efecto en la autorregulación y el aprendizaje la autoevaluación del proceso –guiones– o la autoevaluación del producto final –rúbricas–? Y, segunda, ¿prefieren los alumnos usar rúbricas o guiones? Y si hay una mayor preferencia por uno de estos procedimientos, ¿a qué se debe?

A pesar de que ambas preguntas han sido analizadas por recientes trabajos empíricos se necesita más evidencia como explicaremos a continuación. Hay tres estudios empíricos (Panadero, Alonso-Tapia y Huertas, 2012, 2014; Panadero, Alonso-Tapia y Reche, 2013) que han comparado el uso de rúbricas y guiones en contextos experimentales y entornos reales. Por un lado Panadero et al. (2012) comparando las rúbricas y guiones en un contexto experimental con alumnos de Secundaria encontraron que los guiones aumentaron la autorregulación medida a través de protocolos en voz alta más que las rúbricas, y éstas a su vez más que el grupo control. No obstante, tanto rúbricas como guiones aumentaron por igual el aprendizaje que fue superior en ambos al grupo control. Los otros dos estudios se realizaron con alumnos universitarios en entornos reales. Panadero et al. (2013) encontraron que las rúbricas disminuyeron el uso de estrategias orientadas a la ejecución y a la evitación (Alonso-Tapia,

Panadero y Ruiz, 2014), lo cual es positivo para el aprendizaje, y que el uso de los guiones aumentó el uso de estrategias orientadas al aprendizaje, también positivo. Los participantes en este estudio prefirieron el uso de las rúbricas, lo que los autores atribuyeron a que éstas son más informativas al contener la información relativa a cómo debe ser el resultado final pues contienen ejemplos específicos. No obstante, tanto rúbricas como guiones no aumentaron significativamente el rendimiento pues los alumnos del grupo control obtuvieron calificaciones también altas lo que se atribuyó a un efecto techo en las calificaciones. Por último, Panadero et al. (2014) encontraron que las rúbricas fueron de nuevo preferidas por los alumnos, disminuyeron el uso de estrategias orientadas a la ejecución y a la evitación, y aumentaron el rendimiento académico de los alumnos por encima de los guiones. Como se puede observar, los resultados de los tres estudios no siempre apuntan en la misma dirección. Esto se explica por los diferentes entornos de aplicación –experimental vs. natural–, diferentes medidas de autorregulación –protocolos en voz alta vs. cuestionarios–, diferentes poblaciones –Secundaria vs. Educación Superior–, y tareas diferentes (Panadero, 2011). Por este motivo, seguimos aconsejando que se realicen comparaciones entre guiones y rúbricas atendiendo a la recomendación de que hay que tener en cuenta las características de la intervención para extrapolar los resultados a recomendaciones prácticas.

En conclusión, el uso de rúbricas y guiones presenta efectos positivos que no siempre se encuentran en las intervenciones basadas en la autoevaluación sin criterios de evaluación. Por ello es recomendable el uso de estas herramientas para fomentar la autoevaluación de nuestros alumnos en el aula.

NOTAS

1. Por razones de presentación se describirán en detalle cada una de las categorías en el apartado resultados.

BIBLIOGRAFÍA

- Alonso-Tapia, J. y Panadero, E. (2010). Effect of self-assessment scripts on self-regulation and learning. *Infancia y Aprendizaje*, 33 (3), 385-397.
- Alonso-Tapia, J., Panadero, E. y Ruiz, M.A. (2014). Development and validity of the Emotion and Motivation Self-regulation Questionnaire (EMSR-Q): Intertwined effects between self-regulation styles and classroom motivational climate. *Spanish Journal of Psychology*, 17.
- Andrade, H. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. J. Andrade y G. J. Cizek (Eds.), *Handbook of Formative assessment*. New York: Routledge.
- Andrade, H., Buff, C., Terry, J., Erano, M. y Paolino, S. (2009). Assessment-driven improvements in middle school students' writing. *Middle School Journal*, 40 (4), 4-12.
- Andrade, H. y Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*, 10 (3), 1-11.
<http://pareonline.net/getvn.asp?v=10yn=3>
- Andrade, H., Du, Y. y Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. *Assessment in Education: Principles, Policy & Practice*, 17 (2), 199-214. doi: 10.1080/09695941003696172
- Andrade, H., Du, Y. y Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practices*, 27 (2).
- Andrade, H. y Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice*, 48 (1), 12-19.
- Andrade, H., Wang, X.L., Du, Y. y Akawi, R.L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *Journal of Educational Research*, 102 (4), 287-301.
- Atkinson, R.K., Renkl, A. y Merrill, M.M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology*, 95 (4), 774-783. doi: 10.1037/0022-0663.95.4.774
- Bannert, M. (2009). Promoting self-regulated learning through prompts. *Zeitschrift Fur Pädagogische Psychologie*, 23 (2), 139-145. doi: 10.1024/1010-0652.23.2.139
- Bannert, M., Hildebrand, M. y Mengelkamp, C. (2009). Effects of a metacognitive support device in learning environments. *Computers in Human Behavior*, 25 (4), 829-835. doi: 10.1016/j.chb.2008.07.002
- Berthold, K., Nückles, M. y Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction*, 17 (5), 564-577. doi: 10.1016/j.learninstruc.2007.09.007
- Black, P. y Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*.
- Boekaerts, M. y Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology-an International Review-psychologie Appliquee-revue Internationale*, 54 (2), 199-231.
- Boud, D. (1995). What is learner self-assessment? En D. Boud (Ed.). *Enhancing learning through self-assessment* (pp. 11-23). New York: RoutledgeFalmer.
- Boud, D. y Falchikov, N. (1989). Quantitative studies of student self-assessment in higher-education: A critical analysis of findings. *Higher Education*, 18 (5), 529-549. doi: 10.1007/BF00138746
- Brown, G.T.L. Glasswell, K. y Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing*, 9 (2), 105-121. doi: 10.1016/j.asw.2004.07.001
- Cooper, H. (2010). *Research synthesis and meta-analysis*. Thousand Oaks, California: SAGE.
- Deci, E. L. y Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.
- Dignath, C. y Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level.

- Metacognition and Learning*, 3, 231-264. doi: 10.1007/s11409-008-9029-x
- Dochy, F. (2006). A guide for writing scholarly articles or reviews for the Educational Research Review. *Educational Research Review*.
<http://www.journals.elsevier.com/educational-research-review/>
- Dochy, F., Segers, M. y Sluijsmans, D. (1999). The use of self-, peer- and co-assessment in higher education. A review. *Studies in Higher Education*, 24 (3), 331-350.
- Dreher, M.J. y Brown, R.F. (1993). Planning prompts and indexed terms in textbook search tasks. *Journal of Educational Psychology*, 85 (4), 662-669.
- Elischberger, H.B., y Roebbers, C.M. (2001). Improving young children's free narratives about an observed event: The effects of nonspecific verbal prompts. *International Journal of Behavioral Development*, 25 (2), 160-166.
- Ertl, B., Fischer, F. y Mandl, H. (2006). Conceptual and socio-cognitive support for collaborative learning in videoconferencing environments. *Computers & Education*, 47, 298-315.
- Falchikov, N. y Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59 (4), 395-430.
- Ge, X. y Land, S.M. (2001, November). *Scaffolding students' problem-solving processes on an ill-structured task using questions prompts and peer interactions*. Paper presented at the National Convention of the Association for Educational Communications and Technology, Atlanta.
- Goodrich Andrade, H. (1996). Understanding rubrics. *Educational Leadership*.
- Goodrich Andrade, H. (2001). The effects of instructional rubrics on learning to write. *Current Issues in Education*, 4 (4). <http://cie.ed.asu.edu/volume4/number4/>
- Goodrich Andrade, H. y Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research*, 97 (1), 21-34.
- Goodrich, H.W. (1996). *Student self-assessment: At the intersection of metacognition and authentic assessment*. (57), ProQuest Information y Learning, US.
- Green, R. y Bowser, M. (2006). Observations from the field: Sharing a literature review rubric. *Journal of Library Administration*, 45 (1), 185-202. doi: 10.1300/J111v45n01_10
- Grofle, C.S. y Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction*, 17 (6), 612-634. doi: 10.1016/j.learninstruc.2007.09.008
- Hafner, O.C., y Hafner, P. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education*, 25 (12), 1509-1528. doi: 10.1080/0950069022000038268
- Jonsson, A. y Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130-144.
- Kitsantas, A., Reiser, R.A. y Doster, J. (2004). Developing self-regulated learners: Goal setting, self-evaluation. *The Journal of Experimental Education*, 72 (4), 269-287.
- Kollar, I., Fischer, F. y Slotta, J.D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction*, 17 (6), 708-721. doi: 10.1016/j.learninstruc.2007.09.021
- Kostons, D., Van Gog, T. y Paas, F. (2009). How do I do? Investigating effects of expertise and performance-process records on self-assessment. *Applied Cognitive Psychology*, 23 (9), 1256-1265. doi: 10.1002/acp.1528
- Kramarski, B. y Dudai, V. (2009). Group-metacognitive support for online inquiry in mathematics with differential self-questioning. *Journal of Educational Computing Research*, 40 (4), 377-404.
- Kramarski, B. y Michalsky, T. (2009). Three metacognitive approaches to training pre-service teachers in different learning phases of technological pedagogical content knowledge. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 15 (5), 465 - 485.
- Kramarski, B. y Michalsky, T. (2010). Preparing preservice teachers for self-regulated learning in the context of technological pedagogical content knowledge. *Learning and Instruction*, 20 (5), 434-447. doi: 10.1016/j.learninstruc.2009.05.003

- Kramarski, B. y Zeichner, O. (2001). Using technology to enhance mathematical reasoning: Effects of feedback and self-regulation learning. *Educational Media International*, 38 (2), 77-82.
- Kruger, J. y Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77 (6), 1121-1134.
- Landrum, R.E. (1999). Student expectations of grade inflation. *Journal of Research and Development in Education*, 32 (2), 124-128.
- Lejk, M. y Wyvill, M. (2001). The effect of the inclusion of self-assessment with peer assessment of contributions to a group project: A quantitative study of secret and agreed assessments. *Assessment & Evaluation In Higher Education*, 26 (6), 551-561.
- McCormick, M.J., Dooley, K.E., Lindner, J.R. y Cummins, R.L. (2007). Perceived growth versus actual growth in executive leadership competencies: An application of the stair-step behaviorally anchored evaluation approach. *Journal of Agricultural Education*, 48 (2), 23-35.
- McDonald, B. y Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice*, 10 (2), 209-220.
- Montague, M. (2007). Self-regulation and mathematics instruction. *Learning Disabilities Research & Practice*, 22 (1), 75-83.
- Nicol, D. y McFarlane-Dick, D. (2006). Formative assessment and self-regulated learning, a model and seven principles of good feedback practice. *Studies in Higher Education*, 31 (2), 199-218. doi: 10.1080/03075070600572090
- Noonan, B. y Duncan, C.R. (2005). Peer and self-assessment in high schools. *Practical Assessment, Research & Evaluation*, 10 (17). <http://pareonline.net/getvn.asp?v=10&n=17>
- Panadero, E. (2011). *Instructional help for self-assessment and self-regulation: Evaluation of the efficacy of self-assessment scripts vs. rubrics*. (Tesis), Departamento de Psicología Básica, Universidad Autónoma de Madrid, España.
- Panadero, E. y Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9 (0), 129-144. doi: <http://dx.doi.org/10.1016/j.edurev.2013.01.002>
- Panadero, E., Alonso-Tapia, J. y Huertas, J.A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, 22 (6), 806-813. doi: 10.1016/j.lindif.2012.04.007
- Panadero, E., Alonso-Tapia, J. y Huertas, J.A. (2014). Rubrics vs. self-assessment scripts: Effects on first year university students' self-regulation and performance. *Infancia y Aprendizaje*, 37 (1).
- Panadero, E., Alonso-Tapia, J. y Reche, E. (2013). Rubrics vs. self-assessment scripts effect on self-regulation, performance and self-efficacy in pre-service teachers. *Studies in Educational Evaluation*, 39 (3), 125-132. doi: 10.1016/j.stueduc.2013.04.001
- Petkov, D. y Petkova, O. (2006). Development of scoring rubrics for IS projects as an assessment tool. *Issues in Informing Science and Information Technology*, 3, 499-510.
- Reddy, Y.M. (2007). Effects of rubrics on enhancement of student learning. *Educate*, 7 (1), 3-17.
- Reitmeier, C.A., y Vrchota, D.A. (2009). Self-assessment of oral communication presentations in food science and nutrition. *Journal of Food Science Education*, 8 (4), 88-92.
- Reynolds-Keefer, L. (2010). Rubric-referenced assessment in teacher preparation: An opportunity to learn by using. *Practical Assessment Research & Evaluation*, 15 (8). <http://pareonline.net/getvn.asp?v=15&n=8>
- Sadler, P.M. y Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11 (1), 1-31.
- Schafer, W.D., Swanson, G., Bené, N. y Newberry, G. (2001). Effects of teacher knowledge of rubrics on student achievement in four content areas. *Applied Measurement in Education*, 14 (2), 151-170.
- Schamber, J.F. y Mahoney, S.L. (2006). Assessing and improving the quality of group critical thinking exhibited in the final projects of

- collaborative learning groups. *The Journal of General Education*, 55 (2), 103-137. doi: 10.1353/jge.2006.0025
- Schworn, S. y Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology*, 99 (2), 285-296. doi: 10.1037/0022-0663.99.2.285
- Simon, M. y Forgette-Giroux, R. (2001). A rubric for scoring postsecondary academic skills. *Practical Assessment Research & Evaluation*, 7 (18).
<http://PAREonline.net/getvn.asp?v=7yn=18>
- Tan, K.H.K. (2012). *Student self-assessment. Assessment, learning and empowerment*. Singapore: Research Publishing.
- Taras, M. (2008). Issues of power and equity in two models of self-assessment. *Teaching in Higher Education*, 13 (1), 81-92. doi: 10.1080/13562510701794076
- Taras, M. (2010). Student self-assessment: Processes and consequences. *Teaching in Higher Education*, 15 (2), 199-209. doi: 10.1080/13562511003620027
- Thillmann, H., Kunsting, J., Wirth, J. y Leutner, D. (2009). Is it merely a question of "What" to prompt or also "When" to prompt? The role of point of presentation time of prompts in self-regulated learning. *Zeitschrift Fur Padagogische Psychologie*, 23 (2), 105-115. doi: 10.1024/1010-0652.23.2.105
- Van den Boom, G., Paas, F., Merriënboer, J.J.G. y Van Gog, T. (2004). Reflection prompts and tutor feedback in a web-based learning environment: Effects on students' self-regulated learning competence. *Computers in Human Behavior*, 20, 551-567.
- Ward, M., Gruppen, L. y Regehr, G. (2002). Measuring self-assessment: Current state of the art. *Advances in Health Sciences Education*, 7 (1), 63-81.
- Winne, P.H., y Hadwin, A. F. (1998). Studying as self-regulated engagement in learning. In D. Hacker, J. Dunlosky y A. Graesser (Eds.). *Metacognition in educational theory and practice* (pp. 277-304). Hillsdale, NJ: Erlbaum.
- Winne, P.H., y Jamieson-Noel, D. (2003). Self-regulating studying by objectives for learning: Students' reports compared to a model. *Contemporary Educational Psychology*, 28 (3), 259-276.
- Zimmerman, B.J. (2000). Attaining self-regulation: A social cognitive perspective. En M. Boekaerts, P.R. Pintrich y M. Zeidner (Eds.). *Handbook of self-regulation* (pp. 13-40). San Diego, California: Academic Press