

Evaluación de estrategias metacognitivas: aplicación de métodos online

María Consuelo Sáiz-Manzanares* y Miguel Ángel Queiruga-Dios

Universidad de Burgos

Resumen: Los conocimientos con un grado alto de abstracción tienen dificultades de comprensión para los alumnos. Una metodología instruccional efectiva es el uso del aprendizaje autorregulado (AAR), ya que facilita el empleo de las estrategias metacognitivas (EM) y éstas explican el 40% de la varianza de los resultados de aprendizaje. En la medición de las mismas los métodos de evaluación *on-line* se han mostrado más eficaces que los *off-line*. Los objetivos de este estudio fueron: 1) conocer si existía relación entre las EM registradas con métodos *on-line* vs. métodos *off-line* en el aprendizaje de conceptos de Física y 2) comprobar si la calidad de las EM (orientación, planificación, evaluación y elaboración) depende del tipo de contenido. A lo largo de un curso académico se analizaron 499 protocolos de pensar en voz alta en estudiantes de Educación Secundaria en la asignatura de Física. Se hallaron correlaciones bajas entre las EM (medidas con métodos *on-line*) y las EM de auto-evaluación (medidas con métodos *off-line*). También, se encontraron diferencias significativas en el uso de las EM atendiendo al tipo de contenido. Por ello, la utilización de técnicas de evaluación *on-line* parece facilitar al profesorado el análisis de la calidad de los resultados de aprendizaje.

Palabras clave: Protocolos de pensar en voz alta, Auto-regulación del aprendizaje, Métodos de evaluación en línea, Análisis de textos, Educación Secundaria.

Metacognitive skills analysis: an application of online assessment method

Abstract: Great difficulties have always been associated with learning highly abstract concepts. One effective teaching method is Self-Regulated Learning (SRL), because it facilitates the use of Metacognitive Skills (MS), which have been found to explain 40% of the variance in learning outcomes. Also in the MS measurement is more effective to use *on-line* methods than *off-line* methods. The aims of the study were: 1) to establish whether a relationship exists between MS registered with *on-line* methods to MS registered with *off-line* methods; and 2) to verify whether the quality of each MS (orientation, planning, evaluation and elaboration) depends on the contents of the thematic units. 499 think-aloud protocols in Secondary Education students over one academic year are analyzed in this paper. Relationships between reorientation and planning MS (measures with *on-line* methods) and self-assessment MS (measures with *off-line* methods) were found. Also, we found significant differences in MS use. Thus, the use of online assessment techniques would appear to be of immense assistance to teachers wishing to analyze the quality of the learning outcomes of their students.

Keywords: Think-aloud protocols, Self-regulated learning, *On-line* methods, Text mining, Secondary Education.

Tradicionalmente, la enseñanza de conceptos de Física y de Matemáticas y su aplicación a la resolución de tareas ha presentado dificultades de comprensión

para los estudiantes en distintas etapas del sistema educativo, especialmente en Educación Secundaria y en Universidad. Una posible explicación es el alto componente de abstracción que la comprensión de dichos conceptos exige (Sáiz y Bol, 2015). Además, se añade, en algunos casos, la adquisición de conceptos previos erróneos por parte de los aprendices que dificultan una construcción

Recibido: 14/10/2017 - Aceptado: 5/12/2017 - Avance online: 22/12/2017

*Correspondencia: María Consuelo Sáiz-Manzanares.

Facultad de Ciencias de la Salud. C/ Comendadores s/n.

C.P.: 09001, Burgos, España.

E-mail: mcsmanzanares@ubu.es

correcta del nuevo concepto (Pozo, 2007; Sáiz y Pérez, 2016). En el informe PISA (*Programme for International Student Assessment*) de 2015 (OECD, 2016) se hace referencia a la necesidad de que los estudiantes adquieran contenidos conceptuales y procedimentales que les permitan desarrollar aprendizajes reflexivos, ya que los resultados hallados en el uso de estas habilidades han sido bajos.

Por ello, es preciso plantearse la pregunta de qué tipo de conocimientos y de procesos de pensamiento son necesarios para el buen desarrollo del aprendizaje en los dominios científicos y cómo se puede mejorar la instrucción de los mismos (Reif, 2008). Recientes investigaciones (Azevedo, Johnson, Chauncey, y Graesser, 2011) indican que la utilización de metodologías docentes basadas en AAR (Aprendizaje Autorregulado) junto con la microevaluación de las respuestas de aprendizaje de los estudiantes (Sáiz y Montero, 2015) pueden mejorar las respuestas de aprendizaje.

En primer lugar, la instrucción basada en el AAR guía la enseñanza basada en el *feedback* orientado a procesos (Sáiz y Marticorena, 2016). Además, permite a los estudiantes la evaluación continuada de su proceso de aprendizaje (Sáiz y Montero, 2015). Este tipo de *feedback* es el nivel más efectivo de entre todos los posibles, ya que incrementa el aprendizaje reflexivo, auto-regulado (de Kock, 2016; Norman y Furnes, 2016), la auto-evaluación y por ende la consecución de aprendizajes más profundos (Hattie y Gan, 2011; Sáiz, Marticorena, García y Díez, 2017). En último término, todo ello incrementa en los aprendices la motivación hacia el objeto de aprendizaje (Narciss, 2008; Schunk, Meece, y Pintrich, 2012; Zimmerman y Schunk, 2011).

Siguiendo a Zimmerman (1990), el AAR

tiene que partir de preguntas que ayuden al docente a estructurar adecuadamente las materias, con el objetivo final de facilitar el desarrollo de los subprocesos en la auto-regulación (ver Tabla 1). Este modelo instruccional ha servido de punto de partida a muchos autores para la elaboración de propuestas instruccionales basadas en la auto-regulación metacognitiva (Veenman, 2011). La instrucción basada en AAR se considera un buen predictor del aprendizaje eficaz (Hernández y Camargo, 2017).

Otro elemento referencial es la evaluación durante el AAR. En ella, se puede diferenciar el proceso de auto-evaluación que el sujeto hace de su propio proceso de aprendizaje y la evaluación que de dicho proceso efectúa el docente. Un procedimiento que se ha mostrado muy útil en ambos casos ha sido el empleo de rúbricas (Hattie y Timperley, 2007). Si bien, para que estas sean efectivas, tienen que tener los criterios de evaluación correctamente definidos. Un elemento de ayuda en la redacción de dichos criterios es la utilización de la taxonomía de Bloom (Panadero y Jonsson, 2013). Además, se pueden utilizar otros métodos de evaluación en entornos de AAR. De acuerdo con Schraw (2010) se diferencian dos métodos, los métodos *on-line* (e.g., protocolos de pensar en voz alta y juicios de calibración del aprendizaje) y métodos *off-line* (e.g., cuestionarios o entrevistas). Los primeros se emplean durante los procesos de resolución de tareas, éstos registran las verbalizaciones de los aprendices durante la resolución, o bien recogen los registros de las ejecuciones en el ordenador, estas unidades de información son los *logs-line* (Von Davier, Hao, Liu, y Kyllonen, 2017). Los segundos se pueden administrar antes o después de la resolución de una tarea o de un problema.

Tabla 1
Modelo de auto-regulación propuesto por Zimmerman (1990)

Cuestiones de aprendizaje	Subprocesos de auto-regulación
Por qué (Why)	Auto-eficacia y auto-metas.
Cómo (How)	Estrategias de uso y rutinas de desarrollo.
Cuándo (When)	Tiempo de la instrucción.
Qué (What)	Auto-observación, auto-juicio y auto-reflexión.
Dónde (Where)	Contexto de desarrollo de la instrucción.
Con quién (With Whom)	Estrategias de ayuda.

Los segundos son cuestionarios o escalas que contienen ítems referidos a la frecuencia de utilización de las estrategias procedimentales durante la resolución de problemas. La ventaja de estos últimos, es que se pueden administrar de forma colectiva a muestras amplias de sujetos frente a los primeros en los que su aplicación es individual lo que dificulta el trabajo con muestras amplias. Además, la validez en unos y otros procedimientos es diferente. En los métodos *off-line* se evalúa el recuerdo que el sujeto tiene de cómo resuelve problemas. Esta información puede estar sesgada, ya que hace uso de la Memoria a Largo Plazo (MLP), por lo que el recuerdo puede estar distorsionado debido a la influencia de variables motivacionales relacionadas con la historia de aprendizaje anterior del aprendiz (Sáiz y Payo, 2012; Sáiz y Román, 2011; Veenman, 2011). De otro lado, los métodos *on-line* se aplican durante la ejecución de la tarea, lo que incrementa la validez ecológica ya que los datos se registran en tiempo real. No obstante, tiene el inconveniente de que para que el registro sea referencial se precisa por parte de los aprendices un buen nivel de expresión verbal de los pensamientos. En síntesis, ambos métodos de evaluación difieren en la forma de aplicación y en la validez externa (Veenman, 2011). Además, los resultados hallados en ellos no se pueden equiparar, de hecho se han encontrado coeficientes de correlación muy bajos ($r = .15$) (Veenman et al., 2014). Además, recientes investigaciones han hallado que la aplicación de técnicas de evaluación *on-line* como el análisis de protocolos de pensar en voz alta, facilita la detección de las fortalezas y debilidades de los estudiantes durante el proceso de aprendizaje (Siddiq y Scherer, 2017).

Así pues, la evaluación de las EM que el sujeto emplea en los procesos de resolución de problemas se puede realizar utilizando métodos de evaluación *on-line* u *off-line*. Los primeros no han sido aplicados con mucha frecuencia, ya que precisaban, hasta hace unos años, de procedimientos de registro (grabación de los diálogos de pensar en voz alta) y análisis (categorización de los registros

y posterior análisis e interpretación de los datos) muy complejos (Van der Stel y Veenman, 2014; Veenman, Bavelaar, De Wolf, y Van Haaren, 2014), por lo que tradicionalmente se han utilizado instrumentos *off-line*, es decir, cuestionarios o test, ya que estos permitían un registro más sencillo y un procesamiento más rápido de los resultados. No obstante, en la actualidad, el uso de recursos de aprendizaje automático o Machine Learning facilita tanto el registro como la categorización y posterior análisis de la información recogida. Estos recursos son *softwares* específicos (e.g., dictado de texto o *Speech to Text*) que agilizan el registro de protocolos de pensar en voz alta y su posterior análisis. Un ejemplo son *softwares* como NVivo, ATLAS.ti, MAXQDA, SPAD, KNIME entre otros. Así como, la utilización de paquetes específicos para el análisis de datos como el programa R que contienen técnicas para el análisis de textos (*text mining*), las librerías como WEKA (*Waikato Environment for Knowledge Analysis*), Scikit-Learn en Python, LibreQDA, QDA Miner y VisualText entre otros. Posteriormente, la información recogida se analiza a través de la aplicación de técnicas de minería de datos (Educational Data Mining-EDM-) entre las que se encuentran las técnicas de minería de textos (*text mining*) (Costa, Reis, Neri de Sousa, Moreira, y Lamas, 2018). Dichas técnicas, hacen referencia al procesamiento de grandes volúmenes de datos con la aplicación de técnicas de Inteligencia Artificial (Sáiz, et al., 2017).

En resumen, es preciso evaluar el uso de las EM de los aprendices durante los procesos de aprendizaje, ya que la utilización de dichas estrategias explica el 40% de la varianza de los resultados de aprendizaje (Rodríguez, Piñeiro, Regueiro, Estevez, y Val, 2017; Veenman et al., 2014). Las EM se pueden categorizar en EM de: orientación, planificación, evaluación y elaboración. En la Tabla 2 se presenta una relación de los tipos de EM, los indicadores para su evaluación y los métodos más adecuados en su registro. Asimismo, la utilización de AAR va a facilitar la adquisición, aplicación y calidad (la cual hace referencia a la profundidad en la utilización de las EM por

Tabla 2
Indicadores de evaluación de la calidad de las estrategias metacognitivas en métodos de evaluación on-line,
adaptado de Sáiz y Montero (2015)

Métodos de evaluación de las habilidades metacognitivas	Estrategias metacognitivas	Indicadores de evaluación
Métodos <i>on-line</i> cuantitativos (analiza la frecuencia de aparición de los indicadores de actividades metacognitivas).	Orientación	<ul style="list-style-type: none"> - Activa los conocimientos previos relacionados con la tarea. - Lee adecuadamente el enunciado del problema. - Efectúa un esquema para la representación del problema. - Activa conocimiento previo de las metas. - Sitúa qué es lo que da el problema y qué es lo que pide con el fin de predecir una respuesta.
	Planificación	<ul style="list-style-type: none"> - Expresa las estrategias de resolución mientras analiza el problema. - Señala los pasos de resolución uno a uno.
	Evaluación	<ul style="list-style-type: none"> - Observa y detecta de errores. - Compara las sentencias del problema con las respuestas dadas. - Corrige los errores.
	Elaboración	<ul style="list-style-type: none"> - Realiza paráfrasis y esquema de conclusiones.
Métodos <i>on-line</i> cualitativos (analiza la profundidad de las actividades metacognitivas en una escala de 0, ninguna, a 4, máxima).	Orientación	Instrumentos de observación: protocolos de pensar en voz alta o los registros de log-line en el ordenador.
	Planificación	
	Evaluación	
	Elaboración	

parte de los alumnos) de las EM. La calidad de las EM sólo se puede medir a través del empleo de métodos *on-line*, ya que los *off-line* sólo miden la frecuencia en el uso de las estrategias.

Otro aspecto importante relacionado con el empleo de las EM, es la dificultad del objeto de aprendizaje. Para analizar el uso de las EM en la adquisición los contenidos curriculares se tienen que utilizar escalas de gradación de dificultad creciente que contengan tanto indicadores cuantitativos como cualitativos

(Panadero y Jonsson, 2013; Sáiz y Bol, 2015).

Partiendo de las investigaciones anteriormente señaladas se plantearon dos objetivos de investigación: el primero, comprobar si existía relación entre las estrategias metacognitivas registradas con métodos *on-line* y las registradas con métodos *off-line* en el aprendizaje de conceptos de Física; y el segundo estudiar si la calidad de las respuestas metacognitivas (orientación, planificación, evaluación y elaboración) dependía del tipo de contenido

curricular trabajado. Para ello se formularon las siguientes hipótesis de investigación: H1: ¿Existirá relación entre las estrategias metacognitivas registradas con métodos *on-line* y las registradas con métodos *off-line* en el aprendizaje de conceptos de Física?; y H2: ¿la calidad de las respuestas metacognitivas (orientación, planificación, evaluación y elaboración) dependerá del tipo de contenido curricular trabajado?

MÉTODO

DISEÑO

Se utilizó un diseño descriptivo-correlacional.

PARTICIPANTES

Se trabajó con una muestra de 10 alumnos de 4º curso de Educación Secundaria Obligatoria (ESO) escolarizados en un instituto de la zona centro de una ciudad del norte de España. El nivel socio-económico de las familias era medio-alto. Para la elección del centro y de la muestra se utilizó un muestreo incidental o de conveniencia. El criterio de exclusión fue que los estudiantes no presentasen necesidades educativas especiales. Los datos descriptivos de la muestra atendiendo a las variables edad y sexo se presentan en la Tabla 3.

Asimismo, se analizaron 499 protocolos de pensar en voz alta (verblizaciones estos los alumnos y del docente durante la instrucción curricular) a lo largo de un curso académico.

INSTRUMENTOS

Escalas de Estrategias de Aprendizaje (ACRAr; Román y Poggioli, 2013). Este instrumento está altamente contrastado en distintas poblaciones de habla hispana (Camarero, Martín, y Herrero, 2000; Carbonero, Román, y Ferrer, 2013) e identifica 32 estrategias en distintos momentos del procesamiento de la información. ACRAr tiene altos niveles de fiabilidad y de validez que se sitúan en un intervalo de $\alpha = .78$ a $\alpha = .92$ y $r = .85$ $r = .88$, respectivamente. En este estudio solo se aplicó la Escala de Estrategias Metacognitivas (EEM), en la que se obtuvo un $\alpha = .75$.

Protocolo para el Análisis de la Calidad de las Estrategias Metacognitivas (PACEM; Van der Stel y Veenman, 2014). Es una escala que valora las EM de orientación, planificación, evaluación y elaboración en relación al grado de calidad observado en su utilización. Las respuestas se miden en un gradiente tipo Likert de 1 a 4, desde la ausencia de la utilización de la estrategia metacognitiva (de orientación, planificación, evaluación o elaboración) hasta el empleo sistemático de la misma en la resolución de las tareas. Los datos obtenidos en el análisis de protocolos se estudiaron por dos evaluadores expertos en el análisis de protocolos de pensar en voz alta. Además, se hallaron indicadores altos de fiabilidad para el instrumento en general $\alpha = .84$, y para cada una de las EM: orientación ($\alpha = .78$), planificación ($\alpha = .77$), evaluación ($\alpha = .79$) y elaboración ($\alpha = .84$).

Tabla 3
Estadísticos descriptivos en la variable edad en las distintas muestras

N	Hombres			Mujeres		
	<i>n</i>	<i>M</i> _{edad}	<i>DT</i> _{edad}	<i>n</i>	<i>M</i> _{edad}	<i>DT</i> _{edad}
10	6	17.17	.41	4	17	1.16

Nota. *M*_{edad} = Media de edad; *DT*_{edad} = Desviación Típica

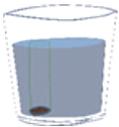
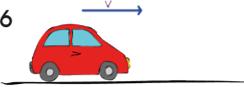
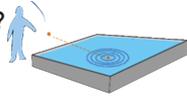
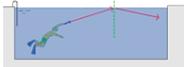
<p>Cada uno de vosotros puede crear su propia unidad de medida... y múltiplos y submúltiplos de esta. Ejemplo tomado de la Unidad 1</p> 	<p>¿Qué opina cada uno de los observadores del movimiento y trayectoria de la moneda? Ejemplo tomado de la Unidad 2</p> 
<p>La Tierra atrae al paracaidista con la misma fuerza con que el paracaidista la atrae a ella. Ejemplo tomado de la Unidad 3</p> 	<p>Representar las fuerzas mediante vectores nos simplifica el problema. Ejemplo tomado de la Unidad 4</p> 
<p>¿Qué soportaría más peso de agua, la moneda o el fondo del vaso?, ¿y presión? Ejemplo tomado de la Unidad 5</p> 	<p>El trabajo realizado por las fuerzas de rozamiento o por la fuerza del motor, modifica la energía del coche. Ejemplo tomado de la Unidad 6</p> 
<p>Una lupa potentísima (imás bien un microscopio!) nos permitiría ver la estructura de la materia. Ejemplo tomado de la Unidad 7</p> 	<p>¿En cuántas dimensiones se propaga la perturbación producida sobre la superficie de agua? Ejemplo tomado de la Unidad 8</p> 
<p>¿Por qué no suena igual la misma nota musical interpretada por distintos instrumentos? Ejemplo tomado de la Unidad 9</p> 	<p>A partir de un cierto ángulo de incidencia, ino se produce refracción! Ejemplo tomado de la Unidad 10</p> 
<p>Figura 1: Ejemplos del uso de las imágenes y de las preguntas de reflexión para facilitar el AAR (tomado de Queiruga, 2009)</p>	

Tabla 4
Relación de contenidos por unidad y tema en del Programa de instrucción en conceptos de Física

Unidades	Temas
Unidad 1	Tema 1. El método científico.
Unidad 2	Tema 2. Las magnitudes físicas.
Unidad 3	Tema 3. El movimiento.
Unidad 4	Tema 4. La Aceleración y el movimiento.
Unidad 5	Tema 5. Los tipos de cuerpos.
Unidad 6	Tema 6. Las fuerzas y el movimiento.
Unidad 7	Tema 7. Las partículas.
Unidad 8	Tema 8. Las fuerzas y la posición.
Unidad 9	Tema 9. La presión.
Unidad 10	Tema 10. La atmósfera.
Unidad 11	Tema 11. La energía
Unidad 12	Tema 12. El trabajo.
Unidad 13	Tema 13. La temperatura.
Unidad 14	Tema 14. El calor.
Unidad 15	Tema 15. El Movimiento y los fenómenos ondulatorios.
Unidad 16	Tema 16. Leyes que explican los movimientos.
Unidad 17	Tema 17. Sonido y sus propiedades.
Unidad 18	Tema 18. Unidades de medida del sonido.
Unidad 19	Tema 19. La Luz.
Unidad 20	Tema 20. El color.

PROCEDIMIENTO

Se obtuvo autorización por parte de la Comisión de Ética del Centro en el que se llevó a cabo el estudio y el consentimiento informado de todos los participantes. Las familias tenían un nivel socio-económico medio-alto. La elección de la muestra se efectuó utilizando un muestreo de conveniencia.

Los estudiantes, en las primeras semanas del curso cumplieron la EEM de ACRAr. La instrucción se efectuó a lo largo de 28 semanas utilizando metodología de AAR. Para ello, se aplicó el *Programa de aprendizaje autorregulado en la asignatura de Física para 4º de ESO* (Queiruga, 2009). Dicho programa consta de 10 unidades sobre conceptos de Física correspondientes al currículo de ESO (ver Tabla 4) en las que se aplica una instrucción basada en AAR. Esta metodología consiste en realizar preguntas a los estudiantes sobre los conceptos a adquirir que se apoyan en imágenes cuidadosamente diseñadas para facilitar la verbalización de posibles respuestas de reflexión sobre el constructo objeto de aprendizaje. Estas respuestas fueron modeladas y moldeadas por parte del docente potenciando la construcción correcta del concepto. Las unidades temáticas se estructuraron en los siguientes apartados: 1) *Objetivos* (competencias a desarrollar en

los estudiantes); 2) *Criterios de evaluación* (sobre los contenidos conceptuales y procedimentales trabajado en cada unidad temática; se medían antes del inicio del tema y después); 3) *Materiales* (recursos necesarios para trabajar la unidad); 4) *Explicación de los contenidos de la unidad temática*, que se apoyaba en dibujos cuidadosamente elaborados y en preguntas de reflexión que facilitaban el AAR, un ejemplo se presenta en la Figura 1; 5) *Actividades de generalización* (actividades semejantes a las trabajadas en la unidad pero con un grado mayor de dificultad).

Además, las interacciones entre el profesor y los estudiantes en cada sesión se grabaron en audio (se registraron 499 protocolos). Posteriormente, se realizó la transcripción de las grabaciones para lo que se utilizó el Programa Atlas ti v.7. Seguidamente para analizar los protocolos de pensar en voz alta se aplicó PACEM de Van der Stel y Veenman (2014). Los registros recogidos en los protocolos se estudiaron por separado por dos evaluadores que aplicaron el PACEM. Posteriormente, se comparó los resultados de la evaluación de los dos evaluadores y se realizó un tercer protocolo de consenso. Para hallar la fiabilidad entre evaluadores en los protocolos analizados se utilizó el coeficiente de contingencia de Pearson aplicando la

Tabla 5
Número de intervenciones registradas por unidad y por tema.

Unidad	Tema			Total	Temas posibles
	1	2	3		
1	15	12	0	27	2
2	34	11	10	55	3
3	0	71	0	71	2
4	26	28	0	54	2
5	17	17	0	34	2
6	42	14	0	56	2
7	46	0	0	46	1
8	65	0	0	65	1
9	41	0	0	41	1
10	50	0	0	50	1
Total	336	153	10	499	2

fórmula siguiente $C = \sqrt{\frac{X^2}{N + X^2}}$ (cf. Meijer et al 2006 citado por Schellings, van Hout-Wolters, Veenman y Meijer, 2013, p. 974). El coeficiente de contingencia entre el investigador y el asistente fue de $r = .96$, por lo que la fiabilidad inter-evaluadores se consideró alta.

Como ya se ha indicado, el programa instruccional incluía 10 unidades temáticas. Asimismo, cada unidad dependiendo de su contenido contenía 1, 2 o 3 temas, registrándose un total de 20. Además, se efectuó una transcripción por tema, la duración media de las grabaciones fue de 38 minutos, computándose 760 minutos de grabación (13 horas). La Media (M) en minutos por registro fue de 33.42 y la Desviación Típica (DT) de 4.80. En cada diálogo se registró la intervención de los alumnos y del profesor. En la Tabla 5, se presenta el número de intervenciones computadas por unidad y por tema. El total de registros fue de 499, la media de actividades analizadas por protocolo de 31.81 y la desviación típica de 19.81.

ANÁLISIS DE DATOS

Se utilizaron estadísticos descriptivos. El coeficiente de correlación tau de Kendall (τ), ya que es la mejor alternativa no paramétrica al coeficiente r de Pearson cuando se trabaja con muestras pequeñas. Indicadores de curtosis y asimetría de los datos hallados en los protocolos de pensar en voz alta. En el análisis de los protocolos se empleó estadística

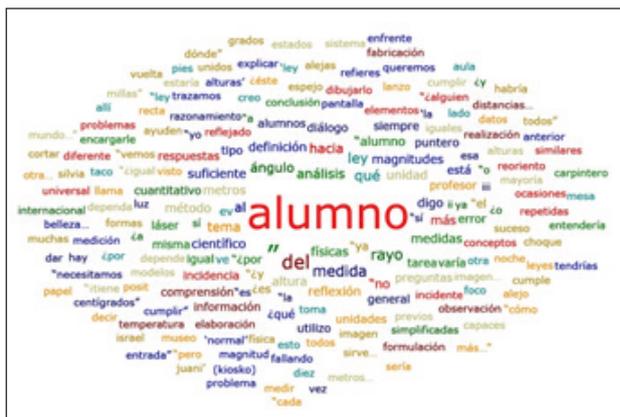


Figura 2: Exploración de las interacciones registradas en el AAR en los diálogos registrados en la Unidad temática 1 en concreto en el Tema 2

paramétrica, ya que el número de protocolos analizados fue de $N = 499$, previamente se habían hallado los indicadores de normalidad (curtosis y asimetría) de la muestra. Por ello, se utilizó un ANOVA de un factor de efectos fijos, valor del efecto y prueba de Bonferroni y un Análisis de Regresión Múltiple (ARM).

El análisis de datos se realizó con el paquete estadístico *Statistical Package for the Social Sciences (SPSS)* en su versión 24 y para la categorización de los protocolos de pensar en voz alta se empleó el Software para análisis de datos cualitativo, gestión y creación de modelos (Atlas.ti) en su versión 7. Este último, es un programa que facilita el análisis de textos, permite la transcripción, codificación y análisis de datos cualitativos. Un ejemplo del análisis de la frecuencia de términos en un diálogo realizado en la instrucción de la unidad temática 1 se puede consultar en la Figura 2.

Tabla 6
Correlaciones de τ entre la calidad de las EM de orientación y de planificación, evaluación y elaboración medidas con métodos online, con las EM medidas con métodos off-line (ACRAR).

	1	2	3	4	5	6	7
Orientación							
Planificación	.91**						
Evaluación	.51**	.51**					
Elaboración	.37**	.40**	.73**				
EM de autoconocimiento	.09	.07	.05	.02			
EM de planificación	-.06	-.04	-.03	-.09	.28**		
EM de auto-evaluación	.11*	.11*	.07	.07	.71**	.29**	

* $p < .05$; ** $p < .01$

RESULTADOS

Respecto de la H1 (¿Existirá relación entre las estrategias metacognitivas registradas con métodos *on-line* y las registradas con métodos *off-line* en el aprendizaje de conceptos de Física?). Sólo se hallaron correlaciones significativas entre las EM de orientación, medidas con métodos *on-line*, y las EM de auto-evaluación medidas con métodos *off-line* ($\tau = .11$) y entre las EM de planificación y las EM de auto-evaluación ($\tau = .11$) (ver Tabla 6).

Posteriormente, para comprobar la H2 (¿la calidad de las respuestas metacognitivas (orientación, planificación, evaluación y elaboración) dependerá del tipo de contenido curricular trabajado?) se realizó previamente un análisis de los valores de asimetría y de curtosis en las respuestas halladas en los protocolos de pensar en voz alta, atendiendo a la clasificación de las EM realizada (orientación, planificación, evaluación y elaboración) por Van der Stel y

Veenman (2014). En asimetría, los valores más altos que $|2.00|$ indican asimetría extrema y los valores más bajos que la distribución es normal. Respecto de los valores de curtosis los valores entre $|8.00|$ y $|20.00|$, sugieren una curtosis extrema (Arias, Verdugo, Navas y Gómez, 2013). Dichos indicadores fueron adecuados (se halló un intervalo entre $|0.30|$ y $|1.88|$ en asimetría $|0.45|$ y $|2.27|$ en curtosis), lo que indica que las distribuciones siguen los parámetros de normalidad (ver Tabla 7).

Seguidamente se realizó un ANOVA de un factor de efectos fijos, se encontraron diferencias significativas entre las distintas unidades temáticas en todas las EM. El valor del efecto más alto se halló en las EM de evaluación ($\eta^2 = .59$) y en las EM de elaboración ($\eta^2 = .67$). Lo que implica que el tipo de unidad temática, es decir el tipo de contenido curricular trabajado, explica la calidad de las EM de evaluación en un 59% y de las de elaboración en un 67% (ver Tabla 8).

Tabla 7

Indicadores de Asimetría y de curtosis en los cuatro indicadores de calidad en las respuestas metacognitivas (orientación, planificación, evaluación y elaboración)

	Mínimo	Máximo	Media	DT	Asimetría	ESM	Curtosis	ESM
Orientación	1.00	4.00	2.48	0.81	-0.46	0.11	-0.53	0.22
Planificación	1.00	4.00	2.38	0.87	-0.30	0.11	-0.88	0.22
Evaluación	1.00	4.00	1.62	0.83	0.96	0.11	-0.45	0.22
Elaboración	1.00	3.00	1.30	0.60	1.88	0.11	2.27	0.22

Nota. DT = Desviación Típica; ESM = Error Estándar de Medida. N = 499.

Tabla 8

ANOVA de un factor de efectos fijos (tipo de unidad temática), y valor del efecto (η^2) y la prueba de comparación de las medias Bonferroni

Calidad de las EM	1	2	3	4	5	6	7	8	9	10	F(10,498)	η^2	Bonferroni
	M(DT)	M(DT)	M(DT)	M(DT)	M(DT)	M(DT)	M(DT)	M(DT)	M(DT)	M(DT)			
Orientación	2.58(1.24)	2.14(.86)	2.32(.69)	2.19(.85)	2.35(.88)	2.54(.81)	2.78(.59)	2.72(.72)	2.85(.53)	2.50(.74)	5.00*	.08	2-7, 2-8, 3-9, 4-7, 4-8, 4-9
Planificación	2.50(1.27)	1.86(.92)	2.13(.86)	2.17(.86)	2.35(.88)	2.25(.79)	2.78(.59)	2.72(.72)	2.83(.54)	2.48(.74)	7.89*	.13	1-2, 10-2, 2-6, 2-7, 2-8, 2-9, 3-7, 3-8, 3-9, 4-7, 4-8, 4-9, 6-9, 7-6
Evaluación	1.04(.20)	1(.00)	1.00(.00)	1.00(.00)	1.00(.00)	1.79(.46)	2.11(.60)	2.32(1.00)	2.83(.54)	2.04(.86)	79.25*	.59	1-6, 1-7, 1-8, 1-9, 1-10, 2-6, 2-7, 2-8, 2-9, 2-10, 3-6, 3-7, 3-8, 3-9, 3-10, 4-6, 4-7, 4-8, 4-9, 4-10, 5-6, 5-7, 5-8, 5-9, 5-10, 6-8, 6-9, 6-10, 7-9, 7-10, 8-9, 8-10, 9-10
Elaboración	1(.00)	1(.00)	1(.00)	1(.00)	1(.00)	1.13(.38)	1.24(.43)	1.51(.50)	2.80(.60)	1.46(.50)	109.89*	.67	1-8, 1-9, 1-10, 2-7, 2-8, 2-9, 2-10, 3-7, 3-8, 3-9, 3-10, 4-7, 4-8, 4-9, 4-10, 5-8, 5-9, 5-10, 6-8, 6-9, 6-10, 7-8, 7-9, 7-10, 8-9, 8-10, 9-10

Nota. EM = Estrategias Metacognitivas; M = Media; DT = Desviación Típica; η^2 = eta cuadrado, valor del efecto. 1 = Unidad Temática 1; 2 = Unidad Temática 2; 3 = Unidad Temática 3; 4 = Unidad Temática 4; 5 = Unidad Temática 5; 6 = Unidad Temática 6; 7 = Unidad Temática 7; 8 = Unidad Temática 8; 9 = Unidad Temática 9; 10 = Unidad Temática 10; N = 499.

Posteriormente, con el fin de comprobar si el contenido curricular de las unidades temáticas era un factor predictor en la calidad de las EM se efectuó ARM. En dicho análisis no se detectaron problemas de multicolinealidad, se entiende que los valores aceptables respecto del factor de inflación de la varianza (VIF) se sitúan entre 1 y 10, como se puede observar en la Tabla 9 los factores son inferiores en todas las variables. Respecto de la Tolerancia (T) (ayuda a identificar si alguna de las variables del modelo es una combinación lineal de las restantes) si el valor de T en alguna variable independiente (VI) es próximo a 0 se puede pensar que esta es una combinación lineal del resto de variables y si el valor de T se aproxima a 1, la variable en cuestión puede reducir parte de la varianza no explicada por el resto de variables. Las variables que presentan una tolerancia pequeña se deben excluir del modelo. En este estudio no se encontraron valores cercanos a 0, por lo que no se eliminó ninguna variable. Asimismo, la correlación parcial más alta se halló entre la EM de evaluación y la variable "tipo de unidad temática" (ver Tabla 9). Además, el Coeficiente de Correlación Múltiple (R^2) fue igual a 0.512, lo que indica que la variable "tipo de unidad temática" explica el 51.2% de la calidad de las EM. Por ello, esta variable se puede considerar como un buen predictor de la calidad de las EM utilizadas por los estudiantes.

DISCUSIÓN

En relación al análisis de los procesos de pensamiento necesarios para la comprensión de conceptos con alto contenido de abstracción, como pueden ser los conceptos trabajados en la materia de Física, se ha encontrado que el uso de técnicas de AAR ha facilitado en los alumnos la reflexión sobre su propio proceso de aprendizaje. Si bien, este tipo de instrucción ha sido más efectiva en las unidades temáticas iniciales, en las que el grado de abstracción de los conceptos era menor (Reif, 2008; Sáiz y Bol, 2015). Una posible explicación es que los conceptos trabajados en estas unidades eran más fáciles de

comprobar experimentalmente que los finales, en los que el componente de abstracción era mayor. Esto sugiere que la dificultad de comprensión del contenido curricular influye en la efectividad del AAR y en la calidad de las EM utilizadas por los estudiantes. Estos resultados están relacionados con los hallados por Pozo (2007) y Sáiz y Montero (2015). Asimismo, la calidad de las EM no se desarrolla de forma uniforme, las estrategias que más respondieron a la instrucción basada en AAR fueron las EM de orientación y de planificación. Dichas estrategias, son las que se utilizan en las fases iniciales de los procesos de resolución de problemas. Las EM de orientación están relacionadas con el análisis de la información y las de planificación con la previsión de los pasos para la resolución de problemas. En cambio, se ha detectado una menor incidencia del AAR en la adquisición de las EM de evaluación y de elaboración, estas son estrategias que se emplean en las fases finales de los procesos de resolución e implican un mayor grado de complejidad metacognitiva que permite la generalización de los aprendizajes (Azevedo et al., 2011). De otro lado, la utilización de métodos de evaluación *on-line* como son los protocolos de pensar en voz alta ha proporcionado mucha información sobre la forma de procesar información de los estudiantes en el aprendizaje de conceptos de Física. Esta labor de microanálisis del proceso de aprendizaje ha sido muy útil para el docente, ya que le ha proporcionado datos acerca de la calidad del aprendizaje y orientado en la utilización de propuestas curriculares más efectivas en cada caso (Sáiz y Marticorena, 2016; Sáiz y Montero, 2015). Asimismo, se ha encontrado una escasa relación entre los resultados de la evaluación de las EM con métodos *on-line* vs. métodos *off-line* aplicado al aprendizaje de conceptos de Física, lo que apoya lo hallado en las investigaciones de Veenman et al. (2014). La información aportada por la aplicación de métodos de evaluación *on-line* incrementa tanto el tipo como la cantidad de información que el docente puede registrar, es lo que Veenman et al. (2014) denominaron calidad en el uso de las EM. Este es un

factor relevante para la práctica docente, lo que previsiblemente potenciará el logro de aprendizajes eficaces (ya que el uso de las EM explica el 40% de la varianza en los resultados de aprendizaje). Además, estas habilidades están directamente relacionadas con el nivel más alto de evaluación en las pruebas del PISA, ya que su uso orienta al alumnado hacia la autonomía en el aprendizaje y a la generalización de los contenidos aprendidos. Por ello, el uso de este tipo de técnicas de evaluación y análisis va a proporcionar al docente más información y de mayor calidad sobre el proceso de aprendizaje de sus alumnos. Además, al registrar todas las interacciones de los estudiantes le va a permitir un seguimiento más personalizado del proceso de enseñanza-aprendizaje en cada alumno. Además, la utilización de herramientas informáticas aplicadas al análisis de protocolos, como por ejemplo Atlas ti, va a facilitar tanto el registro como el estudio de los mismos. De igual modo, la posterior aplicación para el análisis de técnicas de EDM va a mejorar la interpretación del análisis de resultados (Costa et al., 2018; Sáiz, et al., 2017).

No obstante, las limitaciones de este estudio están relacionadas con el tamaño de la muestra, aunque hay que considerar que la utilización de métodos de evaluación *on-line* como es el análisis de protocolos de pensar en voz alta, implica un registro individualizado de todas las respuestas de los estudiantes. Este aspecto, dificulta el trabajo con muestras grandes de sujetos y por ende la generalización de los resultados. No obstante, futuras investigaciones irán dirigidas a ampliar el tamaño muestral, las características de la muestra relacionadas con etapa educativa, la titulación y la materia objeto de estudio. Además, se incrementará el uso de softwares de análisis de textos para el registro y la utilización EDM en el análisis de los datos. De igual modo, futuros estudios se dirigirán a analizar la relación entre el uso de las distintas EM (orientación, planificación, evaluación y elaboración) y en el aprendizaje de diferentes contenidos curriculares. Previsiblemente, los resultados facilitarán la elaboración de

programas personalizados de instrucción lo que mejorará los resultados de aprendizaje en los estudiantes.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Agradecimientos

Queremos agradecer la colaboración de la dirección del centro y de los participantes en este estudio. Así como al Vicerrectorado de Investigación y Transferencia del conocimiento de la Universidad de Burgos por la ayuda para la financiación de la realización de tesis doctorales en 2016.

REFERENCIAS

- Arias, B., Verdugo, M. A., Navas, P., y Gómez, L. E. (2013). Factor structure of the construct of adaptive behavior in children with and without intellectual disability. *International Journal of Clinical and Health, 13*, 155–166. doi: [10.1016/S1697-2600\(13\)70019-X](https://doi.org/10.1016/S1697-2600(13)70019-X)
- Azevedo, R., Johnson, A., Chauncey, A., y Graesser, A. (2011). Use of hypermedia to assess and convey self-regulated learning. En B. J. Zimmerman y D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 102–121). New York: Routledge/Taylor & Francis Group.
- Camarero, F., Martín, F., y Herrero, J. (2000). Estilos y estrategias de aprendizaje en estudiantes universitarios. *Psicothema, 12*, 615–622.
- Carbonero, M. A., Román, J. M., y Ferrer, M. (2013). Programa para “aprender estratégicamente” con estudiantes universitarios: diseño y validación experimental. *Anales de Psicología, 29*, 876–885. doi: [10.6018/analesps.29.3.165671](https://doi.org/10.6018/analesps.29.3.165671)
- Costa, A. P., Reis, L. P., Neri de Sousa, F., Moreira, A., y Lamas, D. (2018). *Computer Supported Qualitative Research*. Switzerland: Springer. doi: [10.1007/978-3-319-43271-7](https://doi.org/10.1007/978-3-319-43271-7)
- de Kock, W. (2016). Speech versus text supported hints in learning to solve word problems. *Computers in Human Behavior, 57*, 300–311. doi: [10.1016/j.chb.2015.11.037](https://doi.org/10.1016/j.chb.2015.11.037)
- Hattie, J., y Gan, M. (2011). Instructions based on feedback. En R. E. Mayer y P. A. Alexander

- (Eds.), *Handbook of Research on Learning and Instruction: Educational Psychology Handbook series* (pp. 249-271). New York: Routledge.
- Hattie, J., y Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112. doi: [10.3102/003465430298487](https://doi.org/10.3102/003465430298487)
- Hernández, A., y Camargo, A. (2017). Autorregulación del aprendizaje en la Educación Superior en Iberoamérica: una revisión sistemática. *Revista Latinoamericana de Psicología*, 49, 146-160. doi: [10.1016/j.rlp.2017.01.001](https://doi.org/10.1016/j.rlp.2017.01.001).
- Narciss, S. (2008). Feedback strategies for interactive learning tasks. En J. M Spector, M. D. Merrill, J. J. G. Van Merriënboer, y M. P. Driscoll (Eds.), *Handbook of Research on Educational Communications and Technology* (3rd ed., pp. 125-144). Mahwah, NJ: Lawrence Erlbaum Associates. doi: [10.4324/9780203880869.ch11](https://doi.org/10.4324/9780203880869.ch11)
- Norman, E., y Furnes, B. (2016). The relationship between metacognitive experiences and learning: Is there a difference between digital and non-digital study media? *Computers in Human Behavior*, 54, 301-309. doi: [10.1016/j.chb.2015.07.043](https://doi.org/10.1016/j.chb.2015.07.043)
- OECD (Organization for Economic Co-Operation and Development). (2016). *Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy*, OECD Publishing, Paris. doi: [10.1787/9789264255425-en](https://doi.org/10.1787/9789264255425-en)
- Panadero, E., y Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129-144. doi: [10.1016/j.edurev.2013.01.002](https://doi.org/10.1016/j.edurev.2013.01.002)
- Pozo, J. I. (2007). *Cambio conceptual y representacional en el aprendizaje y la enseñanza de la ciencia*. Madrid: Antonio Machado.
- Queiruga, M. A. (2009). *¡Física sí!: La física está en lo cotidiano*. A Coruña: Editorial Q.
- Reif, F. (2008). *Applying cognitive science to education: Thinking and learning in scientific and other complex domains*. Cambridge, M. A: The MIT Press.
- Rodríguez, S., Piñeiro, I., Regueiro, B., Estevez, I., y Val, C. (2017). Estrategias cognitivas etapa educativa y rendimiento académico. *Revista de Psicología y Educación*, 12(1), 19-34.
- Román, J. M. y Poggioli, L. (2013). ACRA (r): Escalas de estrategias de aprendizaje. Caracas: Publicaciones UCAB.
- Sáiz, M. C., y Bol, A. (2015). Cómo enseñar y cómo evaluar la resolución de problemas en física: una reflexión sobre la propia práctica. En M. A. Queiruga (Ed.), *Innovación en la enseñanza de las Ciencias: reflexiones, experiencias y buenas prácticas* (pp. 129-146). A Coruña: Editorial Q.
- Sáiz, M. C., y Marticorena, R. (2016). Metacognition. Self-Regulation and Feedback for Object-Oriented Programming Problem-Solving. En J. Benson (Ed.), *Metacognition: Theory, Performance and Current Research* (pp. 43-94). New York: Nova.
- Sáiz, M. C., Marticorena, R., García-Osorio, C. I., y Díez, J. F. (2017). How Do B-Learning and Learning Patterns Influence Learning Outcomes? *Frontiers in Psychology*, 8(745), 1-13. doi: [10.3389/fpsyg.2017.00745](https://doi.org/10.3389/fpsyg.2017.00745)
- Sáiz, M. C., y Montero, E. (2015). Metacognition, Self-regulation and Assessment in Problem-Solving Processes at University. En A. Peña-Ayala (Ed.), *Metacognition: Fundamentals, Applications, and Trends* (Vol. 76 series Intelligent Systems, pp. 107-133). Switzerland: Springer. doi: [10.1007/978-3-319-11062-2_5](https://doi.org/10.1007/978-3-319-11062-2_5)
- Sáiz, M. C., y Payo, R. J. (2012). Auto-percepción del conocimiento en Educación Superior. *Revista Iberoamericana de Psicología y Salud*, 3(2), 159-174. Recuperado de <http://hdl.handle.net/10259/4417>
- Sáiz, M. C., y Pérez, M. I. (2016). Autorregulación y mejora del auto-conocimiento en resolución de problemas. *Psicología desde el Caribe*, 33(1). doi: [10.14482/psdc.33.1.8076](https://doi.org/10.14482/psdc.33.1.8076)
- Sáiz, M. C., y Román, J. M. (2011). Cuatro formas de evaluación en educación superior

- gestionadas desde la tutoría. *Revista de Psicodidáctica*, 16(1), 145-161.
- Schellings, G. L. M., Van Hout-Wolters, B. H. A. M., Veenman, M. V. J., y Meijer, J. (2013). Assessing metacognitive activities: the in-depth comparison of a task-specific questionnaire with think-aloud protocols. *European Journal of Psychology of Education*, 28, 963-990. doi: [10.1007/s10212-012-0149-y](https://doi.org/10.1007/s10212-012-0149-y)
- Schraw, G. (2010). Measuring Self-Regulation in Computer-Based Learning Environments. *Educational Psychologist*, 45(4), 258-266, doi: [10.1080/00461520.2010.515936](https://doi.org/10.1080/00461520.2010.515936).
- Schunk, D. H., Meece, J. R, y Pintrich, P. R (2012). *Motivation in education: Theory, research, and applications*. London: Pearson Higher Ed.
- Siddiq, P., y Scherer, R. (2017). Revealing the processes of students' interaction with a novel collaborative problem solving task: An in-depth analysis of think-aloud protocols. *Computers in Human Behavior*, 76, 509-525. doi. [10.1016/j.chb.2017.08.007](https://doi.org/10.1016/j.chb.2017.08.007)
- Statistical Package for the Social Sciences (SPSS) (Versión 24). [Software de análisis estadístico de datos]. Madrid: IBM.
- Van der Stel, M., y Veenman, M. V. J. (2014). Metacognitive skills and intellectual ability of Young adolescents: a longitudinal study from a developmental perspective. *European Journal of Psychology of Education*, 29(1), 117-137. doi: [10.1007/s10212-013-0190-5](https://doi.org/10.1007/s10212-013-0190-5)
- Veenman, M. V. J. (2011). Alternative assessment of strategy use with self-report instruments: a discussion. *Metacognition and Learning*, 6, 205-21. doi: [10.1007/s11409-011-9080-x](https://doi.org/10.1007/s11409-011-9080-x)
- Veenman, M. V. J., Bavelaar, L., De Wolf, L., y Van Haaren, M. G. P. (2014). The on-line assessment of metacognitive skills in a computerized learning environment. *Learning and Individual Differences*, 29, 123-130. doi: [10.1016/j.lindif.2013.01.003](https://doi.org/10.1016/j.lindif.2013.01.003)
- Von Davier, A. A., Hao, J., Liu, L., y Kyllonen, P. (2017). I Interdisciplinary research agenda in support of assessment of collaborative problem solving: lessons learned from developing a Collaborative Science Assessment Prototype. *Computers in Human Behavior*, 76, 631-640. doi: [10.1016/j.chb.2017.04.059](https://doi.org/10.1016/j.chb.2017.04.059)
- Zimmerman, B.J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25 3-17.
- Zimmerman, B. J., y Schunk, D. H. (2011). Self-regulated learning and performance: An introduction and an overview. En B. J. Zimmerman y D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 1-12). New York: Routledge. doi: [10.4324/9780203839010.ch1](https://doi.org/10.4324/9780203839010.ch1)