

Competencias en el uso de pruebas en argumentación

Beatriz Bravo,¹ Blanca Puig¹ y María Pilar Jiménez-Aleixandre¹

ABSTRACT (Competencies in the use of evidence in argumentation)

The relevance of the competence in the use of scientific evidence in argumentation is discussed in the context of its prominence in PISA and policy. The paper outlines a proposal of two dimensions of the competence: (1) metaknowledge, or knowledge about the use of evidence, including nature of evidence, its role, and criteria for evaluating evidence; and (2) practice of the use of evidence, including the use of evidence in different contexts and the operations involved in it. The different elements in the proposal are discussed drawing from the literature, and particularly from the authors' studies with two tasks set in socio-scientific contexts: evaluation of Watson's claim about differences in intelligence between 'races', and decision about the ecological efficiency of eating secondary versus tertiary carnivore fish. The difficulties found, both in secondary school and in university students, point to the need to pay specific attention to this dimension of argumentation.

KEYWORDS: argumentation, scientific competence, metaknowledge, use of evidence

Las competencias científicas y el uso de pruebas: objetivos del trabajo

Las competencias científicas están situándose en el primer plano de la investigación educativa, sobre todo debido a su utilización desde 1999 por la OCDE como eje en la evaluación internacional PISA (OCDE, 2006) y a la recomendación por la Unión Europea (UE, 2006) de una lista de competencias básicas como aprendizajes imprescindibles. Esta relevancia se refleja en los documentos curriculares de diferentes países, como España. ¿Cuál es el significado de competencia? ¿Representa algo nuevo?

En cuanto al significado, la definición de competencia según PISA es la capacidad de una persona "para reflexionar y aplicar sus conocimientos y experiencias a los problemas que plantea la vida real" (OCDE, 2006, p. 9). En el currículo de Galicia se define como la capacidad de poner en práctica de forma integrada en contextos y situaciones diversos los conocimientos, destrezas y actitudes. En nuestra opinión esta noción no es un mero cambio de término, sino que presenta dos novedades: a) la integración de saberes conceptuales, destrezas y actitudes, y b) el énfasis en la puesta en práctica o *aplicación* de lo aprendido a otros contextos. Esto implica una intención de superar, entre otros, dos problemas del aprendizaje (y la enseñanza) mostrados por la investigación educativa (e.g., Duschl y Grandy, 2008): la fragmentación y las dificultades

en la transferencia o aplicación de lo aprendido a otros contextos.

De las competencias básicas nos interesa la competencia científica, denominada en los currículos de España 'competencia en el conocimiento y la interacción con el mundo físico' (MEC, 2007). Utilizaremos 'competencias científicas' en plural, en referencia a las tres capacidades requeridas para su desarrollo que, como muestra el análisis comparativo de Cañas, Martín y Níeda (2007), son las mismas tanto en PISA (ver OCDE, 2006, p. 27) como en los currículos del MEC:

- 1) Identificar cuestiones científicas (investigables por parte de las ciencias)
- 2) Explicar fenómenos científicamente
- 3) Utilizar pruebas (*evidence*).

Aunque los efectos analíticos y de investigación educativa se traten por separado, está claro que existe una estrecha relación entre identificar cuestiones que pueden ser objeto de estudio por parte de las ciencias, explicar los fenómenos físicos y naturales por medio de modelos científicos (pues en nuestra opinión esta segunda competencia está muy relacionada con la modelización), y utilizar pruebas para, por ejemplo, evaluar modelos científicos.

En este trabajo se aborda la competencia en utilizar pruebas, estrechamente conectada con la evaluación del conocimiento y la argumentación, objeto de nuestro programa de investigación desde 1994. Nuestro objetivo es: proponer una caracterización de las dimensiones o componentes de la competencia de uso de pruebas, basada en la literatura y en las investigaciones de nuestro equipo, particularmente en datos de dos estudios sobre uso de pruebas en contextos científicos de relevancia social sobre ecología y genética.

¹ Departamento Didáctica das Ciencias Experimentais. Av. Xoan XXIII s.n., Universidade de Santiago de Compostela, 15782 Santiago de Compostela, España.

Correos electrónicos: beatriz.bravo@rai.usc.es;
blanca.puig@rai.usc.es; marilarj.aleixandre@usc.es

En el segundo apartado se revisa la literatura sobre estas cuestiones, y en el tercero se discute nuestra caracterización de esta competencia y sus dimensiones, finalizando con las implicaciones educativas y para la investigación.

Argumentación, evaluación del conocimiento y uso de pruebas: marco teórico

Podemos definir la argumentación sobre cuestiones científicas como la evaluación de enunciados de conocimiento a la luz de las pruebas disponibles, lo que requiere la coordinación entre datos y conclusiones. En otro trabajo (Jiménez-Aleixandre, 2008) se discute la centralidad de la evaluación del conocimiento en los procesos de argumentación y en los ambientes diseñados para promoverla. Esta evaluación se basa en las pruebas disponibles en cada momento, pues la aparición de nuevas pruebas o la reinterpretación de las existentes puede llevar a revisar la evaluación. De ahí que el uso de pruebas, parte de los procesos de argumentación, sea una de las competencias científicas básicas.

Utilizamos el término 'prueba' que es la traducción de *evidence*, pues en castellano 'evidencia' significa otra cosa, lo que no necesita ser probado. El *Diccionario de la Real Academia de la Lengua Española (DRAE)* da distintas acepciones de 'prueba', y combinándolas (con algunos añadidos) podemos definirla como: observación, hecho, experimento, señal, muestra o razón con la que se pretende mostrar que un enunciado es cierto o que es falso. En el contexto de la argumentación, las pruebas son entendidas como datos de naturaleza empírica o teórica que sirven para apoyar una conclusión. Desarrollar una perspectiva adecuada sobre la naturaleza de la ciencia requiere entender cómo se generan y validan las pruebas científicas.

Uno de los trabajos de referencia sobre las pruebas es el de Gott y Duggan (1996), quienes proponen como instrumento para evaluar el aprendizaje procedimental en los trabajos prácticos una taxonomía de los 'Conceptos de Prueba' (*Concepts of Evidence*, CoE). Más que a las destrezas procedimentales, estos conceptos se refieren a las decisiones a tomar acerca, por ejemplo, de qué datos seleccionar, cómo seleccionarlos, etc. La taxonomía de Gott y Duggan se revela como una herramienta útil; por ejemplo Schalk (2008) la ha utilizado en el marco de investigaciones realizadas por el alumnado.

Aunque este artículo de Gott y Duggan se refiere a los trabajos prácticos y durante algún tiempo las pruebas han sido tratadas sobre todo en el marco de estos trabajos empíricos, la línea de estudio sobre la argumentación ha hecho recaer la atención sobre el uso de pruebas en el análisis de la información científica, en la evaluación de enunciados, sea o no en el contexto de laboratorio. En otras palabras, esas decisiones son operaciones necesarias no sólo para orientar las investigaciones en el contexto de trabajos prácticos, sino en general en la construcción del conocimiento científico, como muestra el trabajo de Tytler, Duggan y Gott (2001) que analiza la evaluación de enunciados. En nuestra opinión, la propuesta de estos autores, sobre los conceptos o dimensiones

de las pruebas es interesante, pero no es específica del uso de pruebas, sino que incluye otras cuestiones que corresponden a lo que, según PISA (OCDE, 2006), es la competencia de identificar cuestiones científicas, que incluye la capacidad de reconocer los rasgos característicos de una investigación. Podemos ilustrarlo con los ejemplos citados en PISA: qué elementos deben ser comparados, qué variables deberían modificarse o someterse a control, qué información complementaria se requiere o qué medidas han de adoptarse para recoger los datos que hacen al caso (OCDE, 2006, p. 31), que se corresponden con algunos de los CoE de Gott y Duggan.

Una segunda matización que cabría hacer a la propuesta inicial de Gott y Duggan es que sus CoE constituyen más bien requisitos para llevar a cabo una investigación y sólo hacen referencia a las pruebas empíricas, no a las de naturaleza teórica; sin embargo, en trabajos posteriores como el de Tytler *et al.* (2001) se incluyen también éstas. Cabe hacer notar que en el modelo de Toulmin (1958) utilizado como instrumento para analizar las estructuras de los argumentos, las justificaciones pueden corresponder a pruebas de naturaleza teórica, por ejemplo un modelo de transferencia de energía en el ecosistema.

¿Cómo se utilizan las pruebas en el contexto del aula y en la vida diaria? Mientras que en una investigación se elaboran las conclusiones a partir de los datos, en la evaluación de un enunciado científico el camino puede ser inverso. Es a este proceso inverso, de identificación de pruebas que apoyan o refutan un enunciado, al que se enfrenta la ciudadanía, por ejemplo en cuanto a informaciones en los medios de comunicación. En la línea de investigación sobre argumentación acerca de cuestiones socio-científicas distintos trabajos analizan cómo evalúa el alumnado informaciones científicas. Kolstø (2001), en un estudio con alumnado de 16 años sobre los riesgos potenciales de las líneas de alta tensión, examinó la veracidad que atribuían a los enunciados, encontrando cuatro tipos de estrategias que podían combinarse: aceptación, evaluación usando indicadores de 'fiabilidad', aceptación de la autoridad de los investigadores o fuentes citadas, y evaluación de las fuentes en términos de neutralidad y de competencia. Aunque las evaluaciones del alumnado estaban parcialmente basadas en las pruebas empíricas, tuvieron muy en cuenta informaciones superficiales, y Kolstø concluye que es necesario trabajar sobre la fiabilidad de distintas fuentes, en la perspectiva de una enseñanza de las ciencias para la ciudadanía. En un estudio sobre el determinismo biológico (Puig y Jiménez-Aleixandre, en prensa) con alumnado de secundaria y universidad se ponen de manifiesto dificultades para identificar pruebas para refutar o probar unas declaraciones de James Watson afirmando que los negros eran genéticamente menos inteligentes que los blancos. La evaluación de pruebas es necesaria para la resolución de un dilema medioambiental, como la creación de plantas de acuicultura de especies carnívoras en el litoral gallego: Bravo y Jiménez-Aleixandre (en prensa) analizan respuestas a las preguntas ¿qué es más eficiente desde el punto de vista ecológico, alimentarse de espe-

cies en niveles tróficos superiores o inferiores?, ¿en qué condiciones es la acuicultura una solución a la sobreexplotación de recursos pesqueros? Resolver este dilema requiere tanto apelar a datos empíricos, como aplicar los conceptos teóricos relevantes (flujo de energía, pirámide trófica) a una situación nueva. Dicho de otro modo, no sólo es necesario disponer de datos sino también saber utilizarlos en coordinación con la teoría. Estos estudios ilustran las dos dimensiones que, en nuestra opinión, forman parte de esta competencia: por un lado aspectos que se refieren al conocimiento sobre las pruebas, como la fiabilidad de las fuentes, explorada por Kolstø, y por otro, la práctica de usar pruebas, en nuestros propios estudios.

Utilizar pruebas científicas requiere identificar la información relevante, seleccionarla y llegar a conclusiones basadas en pruebas, es decir argumentar (Kuhn, 1992; Osborne *et al.*, 2001). Uno de los objetivos de la enseñanza de las ciencias reflejado en el currículo español (y en los de otros países) es “adoptar aptitudes críticas fundamentadas en el conocimiento para analizar, individualmente o en grupo, cuestiones científicas o tecnológicas” (MEC, 2007, p. 693). Para lograrlo es necesario diseñar ambientes de aprendizaje y tareas que demanden la práctica de estas competencias por parte del alumnado. Desde la investigación en didáctica se ha abordado de distinta forma cómo desarrollar la competencia de usar pruebas: Gott y Duggan (1996), por ejemplo, consideran que la comprensión de las pruebas requiere un cuerpo de conocimiento que debe enseñarse explícitamente; por otra parte, autores como Sandoval y Millwood (2008) centran su análisis en el desarrollo a través de la práctica. Entendemos que ambas dimensiones son necesarias, ya que para poder utilizar de

forma adecuada las pruebas es necesario tanto un conocimiento básico como un contexto adecuado donde a través de la participación discursiva el alumno aprenda a construir argumentos apoyados en ellas.

Una propuesta de dimensiones en la competencia de uso de pruebas

En nuestra opinión pueden distinguirse dos dimensiones en la competencia de uso de pruebas: una relacionada con el metaconocimiento o conocimiento sobre el uso de pruebas, que se correspondería más con los CoE de Gott y Duggan (1996) y otra relacionada con la práctica o desempeño del uso de pruebas. Ambas están interrelacionadas, desarrollándose conjuntamente. Así, en cuanto al metaconocimiento, es necesario entender cuál es el papel de las pruebas para poder integrar aquellas que son relevantes dentro de una explicación, y al mismo tiempo es preciso dominar las diferentes operaciones o procesos implicados en la puesta en práctica del uso de pruebas. Las dimensiones de la propuesta, que se detallan a continuación, en conexión con los trabajos sobre esta cuestión, se representan en la figura 1.

Metaconocimiento sobre uso de pruebas

La primera dimensión puede definirse como el conocimiento sobre el uso de pruebas, y podemos decir que forma parte de lo que sería la comprensión e imagen del alumnado sobre la naturaleza de la ciencia. Para elaborar nuestra propuesta de los distintos aspectos que comprende este conocimiento, nos hemos basado en el esquema propuesto por Schwarz *et al.* (en prensa) para otra competencia, la de uso de modelos, adaptándolo al uso de pruebas. Puede decirse que este cono-

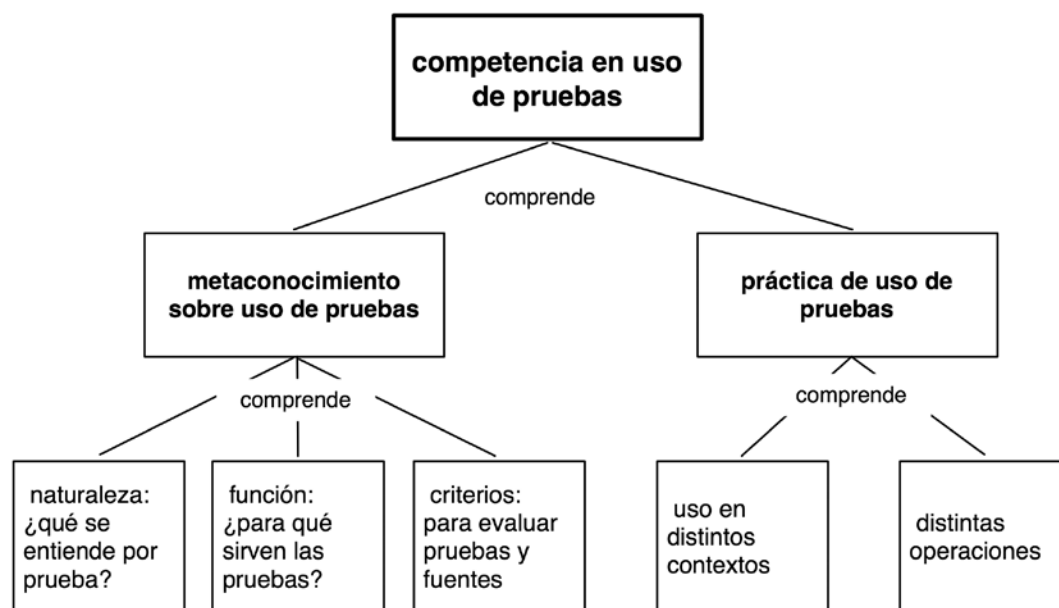


Figura 1. Propuesta de dimensiones en la competencia de uso de pruebas.

cimiento sobre las pruebas y su uso comprende tres aspectos: 1) la naturaleza de las pruebas; 2) su función, y 3) los criterios que se utilizan para evaluarlas.

1) Naturaleza de las pruebas: respondería a la pregunta ¿Qué se entiende por prueba?

En el apartado anterior hemos propuesto una definición, siguiendo el DRAE, como observación, hecho, experimento, señal, muestra o razón, propuesta en la que se incluyen pruebas tanto de naturaleza empírica como teórica. Esta inclusión es importante, puesto que, como muestra el estudio de Sandoval y Millwood (2008), la mayoría de los estudiantes de secundaria preguntados por “la mejor manera de convencer a alguien de alguna cuestión de ciencias”, citaban algún tipo de pruebas empíricas y sólo una pequeña proporción aludían a una explicación de carácter teórico. Otra cuestión relevante es tener en cuenta que en la definición de argumentación nos referimos a las pruebas disponibles, ya que o bien las mismas pruebas pueden ser interpretadas de distintas formas, como ocurrió con los fósiles que son interpretados de una manera según los modelos fijistas y de otra en el modelo de evolución de Darwin y Wallace; o bien pueden aparecer pruebas nuevas que lleven a la modificación de una conclusión, como ha ocurrido en las últimas décadas con el modelo de herencia genética que ha llevado a modificaciones en la propia definición de gen.

2) Función de las pruebas: respondería a la pregunta ¿Para qué sirven las pruebas?

En la definición de prueba del apartado anterior se indica que sirven para mostrar que un enunciado es cierto o que es falso, para apoyar o refutar una conclusión. Podemos añadir su papel en la elección entre teorías, o en la toma de decisión entre distintas opciones. Todo esto se refiere al papel en la justificación, pero hay que tener en cuenta que las pruebas también juegan un importante papel en la persuasión, en convencer a otras personas o a la comunidad científica.

3) Criterios para identificar y evaluar pruebas: respondería a la pregunta ¿qué pruebas son relevantes en la justificación de un determinado enunciado?

En cuanto a los criterios, tendríamos que distinguir por un lado los que se refieren a la evaluación de pruebas, y por otro los que se refieren a la evaluación de fuentes de información. Algunos criterios de evaluación de pruebas serían si la prueba tiene relación con el enunciado (Puig y Jiménez-Aleixandre, en prensa); o, como proponen Kenyon, Kuhn y Reiser (2006) si es específica para la cuestión discutida, y si se origina en datos, no en opiniones.

Práctica o desempeño del uso de pruebas

La segunda dimensión puede definirse como la práctica o desempeño del uso de pruebas, es decir la capacidad para utilizar las pruebas en la evaluación de enunciados, sean teóricos o provengan de resultados empíricos. Proponemos que esta ca-

pacidad comprende dos aspectos, muy interrelacionados entre sí: 1) el uso de pruebas en *distintos contextos*; 2) las operaciones que forman parte de este desempeño.

1) Uso de pruebas en distintos contextos de construcción del conocimiento: este aspecto guarda relación con el papel de las pruebas, que forma parte de la primera dimensión.

Algunos ejemplos, no exhaustivos, de distintos contextos en los que la práctica de evaluación de pruebas presenta algunas diferencias son: a) en la elección entre varias teorías o modelos explicativos de un fenómeno, como pueden ser entre el modelo heliocéntrico y geocéntrico; entre el modelo de origen del relieve por la contracción terrestre o por la tectónica global; entre las teorías que atribuían el cólera y otras enfermedades infecciosas a contagios (contagionistas) y las que los atribuían a miasmas; b) en la elección entre varias opciones o decisiones, por ejemplo entre comer peces que son carnívoros primarios o secundarios, como arenque o sardina, y comer carnívoros terciarios como el salmón, en función de su eficiencia ecológica (Bravo y Jiménez-Aleixandre, en prensa); c) en la evaluación de enunciados de otras personas, como puede ser la afirmación de James Watson sobre las diferencias genéticas en inteligencia entre personas blancas y negras (Puig y Jiménez-Aleixandre, en prensa); d) en la identificación de los supuestos que subyacen a una conclusión determinada, como puede ser, en el caso anterior de Watson, identificar el supuesto de que las diferencias en inteligencia se deben únicamente al genotipo, sin intervención de los factores ambientales.

2) Operaciones que forman parte del desempeño de uso de pruebas: estas operaciones pueden ser diferentes en los distintos contextos de construcción del conocimiento.

Algunas de estas operaciones son, por ejemplo y a título no exhaustivo: evaluar la fiabilidad de las fuentes de información de quien emite el enunciado (experto, libro, docente, medio de comunicación etc.), así en el estudio de Kolstø (2001), citado en el apartado anterior, o en el de Jiménez-Aleixandre, Agraso y Eirexas (2004) sobre afirmaciones contrapuestas de científicos acerca de la marea negra del *Prestige*. En la evaluación de enunciados, como el de James Watson, Puig y Jiménez-Aleixandre (en prensa) proponen al menos tres: identificación del significado del enunciado (que implica comprensión de textos científicos) y que, como muestran estas autoras no siempre tiene lugar; coordinación de enunciados con pruebas, o en otras palabras identificación de qué pruebas son relevantes para apoyar o rechazar el enunciado; e identificación del significado de cada prueba, es decir si apoya o refuta un enunciado determinado. En la evaluación de distintas opciones, como en el trabajo de Bravo y Jiménez-Aleixandre (en prensa), habría que añadir a éstas la articulación de pruebas empíricas (por ejemplo la cantidad de kilos de pequeños pescados necesaria para producir un kilo de salmón) con justificaciones

teóricas, como puede ser el modelo de flujo de energía en un ecosistema, o la pirámide trófica.

La distinción de estas dimensiones tiene propósitos analíticos para guiar tanto el diseño de ambientes de aprendizaje y tareas que favorezcan el desarrollo de esta competencia, como la investigación sobre ella. Distinguir las no significa que se desarrollen por separado, ni que las actividades de enseñanza tengan como objetivo sólo la primera o la segunda, pues están estrechamente relacionadas entre sí.

Puede decirse que algunas actividades para promover el uso de prácticas implican de forma más clara el desarrollo del metaconocimiento, como es el caso en las que demandan implícita o explícitamente la evaluación de las fuentes. Así, en el trabajo de Hogan y Maglienti (2001) donde se propone al alumnado que evalúen las conclusiones de otros estudiantes sobre un problema medioambiental creado por una planta invasora. Los criterios utilizados por el alumnado son comparados con los utilizados por expertos. La diferencia entre ambos es que los expertos utilizan como criterio la existencia de datos empíricos, mientras que las y los estudiantes recurren más a las inferencias personales. En esta línea se sitúa el trabajo de las coautoras (Puig y Jiménez-Aleixandre, en prensa) que requiere por parte del alumnado el análisis de distintos ítems sobre la interacción entre genotipo y ambiente, y su empleo en la evaluación positiva o negativa de la afirmación de Watson sobre las diferencias genéticas de inteligencia entre negros y blancos.

Otras tareas ponen más el acento en la práctica de uso de pruebas, por ejemplo aquellas en las que el alumnado tiene que elaborar una conclusión a partir de una serie de datos, que pueden ser suministrados o bien estar implícitos en la tarea, lo que requiere que sean recuperados del conocimiento escolar. Así el estudio de Kelly y Takao (2002) en el que se analizan los niveles epistémicos en la argumentación de estudiantes universitarios en oceanografía, que disponían de una base de datos reales sobre características y fenómenos tectónicos en diferentes placas y lugares del mundo. Eirexas y Jiménez-Aleixandre (2007) analizan la argumentación de alumnado del último curso de bachillerato (17-18 años) en la elección de un sistema de calefacción que causase el mínimo impacto ambiental posible. El alumnado disponía de un dossier con información de las distintas fuentes de energía, tanto procedente de las propias compañías suministradoras de electricidad, gas, etc., como de estudios comparativos de precios, además del texto de Ciencias de la Tierra y del medio ambiente. Otro ejemplo de trabajo sobre la dimensión práctica es la investigación de la primera autora: la tarea requería que los alumnos seleccionasen datos de un texto sobre las consecuencias de la sobreexplotación pesquera y los relacionasen con sus conocimientos de ecología, aplicando conceptos como pirámide trófica o flujo de energía, para decidir si es más eficiente alimentarse de niveles inferiores o superiores de la cadena trófica (Bravo-Torija y Jiménez-Aleixandre, en prensa).

En algunos trabajos, como el de Sandoval y Millwood (2005) se combinan ambas dimensiones. Estos autores anali-

zan la calidad de los argumentos escritos relacionados con problemas sobre evolución, en cuanto a tres aspectos: la calidad conceptual; la suficiencia de datos, y las referencias retóricas en los argumentos del alumnado. Tanto el primero como el segundo pertenecen fundamentalmente a la dimensión práctica, pero al mismo tiempo el segundo y el tercero se refieren respectivamente a los criterios de evaluación y a la función de persuasión.

Todos los estudios mencionados ponen de manifiesto las dificultades del alumnado para coordinar los datos y conclusiones, las limitaciones de su comprensión del papel de las pruebas y su valor en la argumentación.

A modo de conclusión

Los trabajos de las autoras, en un caso sobre la evaluación de un enunciado acerca de diferencias genéticas en inteligencia entre blancos y negros, y en otro sobre la elección, con criterios de eficiencia ecológica, de si comer salmón o los pescados de los que éste se alimenta, han puesto de manifiesto que, tanto en secundaria como entre alumnado universitario, el uso de pruebas en la construcción de argumentos presenta serias limitaciones. Algunas de estas dificultades pueden relacionarse con los problemas en la comprensión de textos, otras con una insuficiente comprensión de la naturaleza del trabajo científico y, en particular, de la naturaleza y función de las pruebas, y otras con las dificultades para transferir a un contexto lo aprendido en otro (Puig y Jiménez-Aleixandre, en prensa). El análisis de la literatura muestra dificultades similares en otros contextos y tareas. Esto apunta a la necesidad de disponer de herramientas analíticas que permitan, tanto abordar las distintas dimensiones del uso de pruebas en el diseño, como examinar su desarrollo en el aula. Éste es el objetivo de la herramienta aquí presentada, que estamos usando en nuestro trabajo en curso.

Agradecimientos

Trabajo parte del proyecto código SEJ2006-15589-C02-01/EDUC, financiado por el Ministerio de Educación y Ciencia (MEC) de España, y parcialmente con fondos FEDER (Fondo Europeo de Desarrollo Regional). El trabajo de Beatriz Bravo está financiado por una beca FPI del MEC, código BES-2007-15075, y el de Blanca Puig por un convenio con el Consello de Cultura Galega, código 2008/CI78.

Referencias

- Bravo-Torija, B. and Jiménez-Aleixandre, M.P. Is raising salmon sustainable? Use of concepts and evidence about ecology. In: K. Boersma and A. Waarlo (eds.), *Proceedings of the ERI-DOB conference*. Utrecht: University of Utrecht (en prensa).
- Cañas, A., Martín-Díaz, M.J. y Nieda, J., *Competencia en el conocimiento e interacción con el mundo físico* (p. 144), Alianza Editorial: Madrid, 2007.
- Duschl, R.A. and Grandy, R.E. Reconsidering the character and role of inquiry in school science: Framing the debates. In: R.A. Duschl and R.E. Grandy (eds.), *Teaching scientific*

- inquiry. *Recommendations for research and implementation* (pp. 13–7), Sense Publishers: Rotterdam, 2008.
- Eirexas, F. and Jiménez-Aleixandre, M.P. *What does sustainability mean? Critical thinking and environmental concepts about energy by 12th grade students*. Paper presented at the ESERA conference, Malmoe, August, 2007.
- Gott, R. and Duggan, S. Practical work: its role in the understanding of evidence in science, *International Journal of Science Education*, 18(7), 791-806, 1996.
- Hogan, K. and Maglienti, M. Comparing the epistemological underpinnings of students and scientists' reasoning about conclusions, *Journal of Research in Science Teaching*, 38(6), 663-687, 2001.
- Jiménez-Aleixandre, M.P., Designing argumentation learning environments. In: S. Erduran and M.P. Jiménez-Aleixandre (eds.), *Argumentation in science education: perspectives from classroom-based research* (pp. 91-115), Dordrecht: Springer, 2008.
- Jiménez-Aleixandre, M.P., Agraso, M.F. and Eirexas, F., *Scientific authority and empirical data in argument warrants about the Prestige oil spill*. Paper presented at the National Association for Research in Science Teaching annual meeting. Vancouver, April, 2004.
- Kelly, G.J., Regev, J. and Prothero, W., Analysis of lines of reasoning in written argumentation. In: S. Erduran and M.P. Jiménez-Aleixandre (eds.), *Argumentation in science education: perspectives from classroom-based research* (pp. 137-159), Dordrecht: Springer, 2008.
- Kelly, G.J. and Takao, A. Epistemic levels in argument: an analysis of university oceanography students' use of evidence in writing, *Science Education*, 83(3), 115-130, 2002.
- Kenyon, L., Kuhn, L. and Reiser, B.J., Using students' epistemologies of science to guide the practice of argumentation. In: S.A. Barab, K.E. Hay and T.D. Hickey (eds.), *Proceedings of the 7th International Conference of the Learning Sciences* (pp. 321–327), Mahwah, NJ: Lawrence Erlbaum, 2006.
- Kolstø, S.D. 'To trust or not to trust, ...' –pupils' ways of judging information encountered in a socio-scientific issue, *International Journal of Science Education*, 23(9), 877-901, 2001.
- Kuhn, D. Thinking as argument, *Harvard Educational Review*, 62(2), 155-178, 1992.
- MEC, *Real Decreto 1631/2006 por el que se establecen las enseñanzas mínimas correspondientes a la Educación Secundaria Obligatoria*. BOE 5/01/2007, Madrid.
- OCDE, PISA 2006. Marco de la evaluación: *Conocimientos y habilidades en Ciencias, Matemáticas y lectura*. Santillana / Ministerio de Educación y Ciencia: Madrid, 2006. 192 pp.
- Osborne, J., Erduran, S., Simon, S. y Monk, M. Enhancing the quality of argument in school science, *School Science Review*, 82(301), 63-70, 2001.
- Puig, B. and Jiménez-Aleixandre, M.P. What do 9th grade students consider as evidence for or against claims about genetic differences in intelligence between black and white "races"? In: K. Boersma and A. Waarlo (eds.) *Proceedings of the ERIDOB conference*. Utrecht: University of Utrecht, (en prensa).
- Sandoval, W.A. and Millwood, K. The quality of students use of evidence in written explanations, *Cognition and Instruction*, 23(1), 23-55, 2005.
- Sandoval, W. and Millwood, K. What can argumentation tell us about epistemology? In: S. Erduran and M.P. Jiménez-Aleixandre (eds.), *Argumentation in science education: perspectives from classroom-based research*, (pp. 71-90), Dordrecht: Springer, 2008.
- Schalk, H. *The use of concepts of evidence by students in biology investigations: development research in pre-university education*. Paper presented at the ERIDOB conference. Utrecht, 2008.
- Schwarz, C.V., Reiser, B.J., Davis, E.A., Kenyon, L.O., Acher, A., Fortus, D., Hug, B., and Krajcik, J., Designing a learning progression of scientific modeling: Making scientific modeling accessible and meaningful for learners, *Journal of Research in Science Teaching* (en prensa).
- Toulmin, S. *The uses of argument*. Cambridge: Cambridge University Press, 1958.
- Tytler, R., Duggan, S. and Gott, R., Dimensions of evidence, the public understanding of science and science education, *International Journal of Science Education*, 23(8), 815-832, 2001.
- UE, Recomendación del Parlamento Europeo sobre las competencias clave para el aprendizaje permanente. *Diario Oficial de la UE*, 30/12/2006, Bruselas.