

Primeros resultados de una prueba piloto para evaluar el pensamiento histórico de los estudiantes.^{1 2}

por *Jesús Domínguez*

Universidad Nebrija, España

dominguez.castillo.jesus@gmail.com

Laura Arias, Raquel Sánchez, Alejandro Egea

Universidad de Murcia, España

larias@um.es, raqueledu@um.es, alexegea@um.es

Francisco Javier García

Instituto Nacional de Evaluación Educativa, MECD, España

javier.grespo@mecd.es

Recibido: 28/04/2017 - Aceptado: 12/05/2017

Resumen

Este artículo analiza el estudio piloto y primeros resultados de una prueba de historia que ha sido concebida y diseñada como posible modelo para evaluar las competencias de pensamiento histórico del alumnado al finalizar la etapa de secundaria obligatoria. El artículo hace una breve exposición del marco teórico del estudio y explica cómo pueden concretarse en destrezas evaluables los conceptos de pensamiento histórico, como el de empatía. El estudio piloto se ha realizado en dos fases en las que han participado en torno a 800 alumnos de entre 16 y 18 años. Los resultados de la primera fase, con las respuestas de 316 estudiantes analizadas estadísticamente siguiendo la Teoría de Respuesta al Ítem (TRI), nos hacen confiar en que la prueba sea un instrumento adecuado para evaluar dicho pensamiento histórico y para distinguir escalas de niveles de logro en las distintas competencias históricas.

Palabras claves

pensamiento histórico, marco de evaluación, competencias, estudio piloto, análisis TRI.

First results of a pilot test to evaluate students' historical thinking.

Abstract

This article analyzes the pilot study and first results of a test of history that has been conceived and designed as a possible model to assess historical thinking skills of students at the end of compulsory secondary education. The article provides a brief presentation of the assessment framework and explains how historical thinking concepts, such as "empathy", can be developed into measurable skills. The pilot study was conducted in two phases which have involved around 800 students aged between 16 and 18 years old. The results of the first phase, with 316 students' answers that have been analyzed statistically according to the Item Response Theory (IRT), allow us to expect that the test can be an appropriate tool to assess students' historical thinking and to distinguish scales of levels of achievement in the major historical competences.

Keywords

historical thinking, assessment framework, competences, pilot study, IRT analysis

Introducción

Este artículo analiza el estudio piloto y primeros resultados de una prueba de historia que ha sido concebida y diseñada como posible modelo para evaluar las competencias de pensamiento histórico de los estudiantes españoles al terminar la Educación Secundaria Obligatoria (en adelante ESO), con edades comprendidas entre los 16 y 18. Como corresponde a un estudio piloto, el objeto del mismo no ha sido evaluar los conocimientos históricos de los estudiantes que lo han respondido, sino evaluar la prueba y analizar hasta qué punto puede ser un instrumento adecuado para calibrar el pensamiento histórico de los estudiantes.

Siguiendo las pautas que han difundido renombrados estudios internacionales de evaluación como PISA, nuestra prueba no trata de evaluar el conocimiento escolar que los estudiantes han memorizado y, en el mejor de los casos, su capacidad para organizarlo y presentarlo de manera clara y coherente. Por el contrario, la prueba tiene por objetivo evaluar su competencia para utilizar y aplicar ese conocimiento escolar a situaciones o problemas nuevos. En el caso de la historia, para la que no existen pruebas internacionales de las que servirnos como ejemplo, esta competencia podría equivaler, en nuestra opinión, a lo que destacados especialistas en didáctica de la historia (Lee, Shemilt, Rüsen, Seixas) han denominado "pensar históricamente" (*historical thinking*). Además de conocimientos o contenidos, pensar históricamente contemplaría también, según esos autores, entender cómo es el conocimiento del pasado que construyen los historiadores, en qué bases empíricas descansa, cómo explican e interpretan lo ocurrido, e incluso de qué forma tal conocimiento del pasado condiciona nuestra visión del presente y del futuro.

Otro de los rasgos que nuestra prueba toma de los estudios internacionales de evaluación es que no pretende ser un instrumento para evaluar individualmente a cada estudiante, sino poder servir de modelo embrionario para futuras evaluaciones de diagnóstico en historia. Para ello se requieren pruebas muy amplias, con gran número de ítems (hasta 150 en algunos casos), que hacen posible evaluar destrezas diversas en distintos contextos y niveles de dificultad. Esos ítems se distribuyen necesariamente entre los estudiantes, que responden solo a una parte de los mismos. De este modo, tales pruebas ofrecen resultados muy precisos sobre el rendimiento medio, los niveles, progresión, problemas de aprendizaje, etc. del conjunto de los estudiantes. Como tales pueden incluso aplicarse a los estudiantes de un mismo centro, aunque con un margen de error considerable, pero en modo alguno proporcionan datos suficientes para evaluar el rendimiento individual de cada estudiante. Ahí reside su carácter de diagnóstico, toda vez que su valor primordial es el de conocer, analizar y mejorar, nunca calificar, al alumnado.

Finalmente, un estudio como el que presentamos es además necesario para la reflexión sobre el papel de la historia en un currículo obligatorio de orientación competencial. Así lo sugieren las orientaciones curriculares de muchos países, en línea con las recomendaciones de organismos internacionales como la UE o la OCDE. Contribuir a esa reflexión general sobre qué y cómo debemos enseñar la historia es sin duda el objetivo último de este estudio. En efecto, como hemos dicho en otras ocasiones, reflexionar sobre la evaluación trasciende sin duda el ámbito estricto de lo que significa evaluar, medir, etc. el aprendizaje de historia. Nos obliga a considerar también qué debemos priorizar en la enseñanza de la historia. Y es que enseñanza y evaluación son obviamente las dos caras de una misma moneda.

El primer apartado de este trabajo describe el marco teórico de la prueba, muy en consonancia con los estudios de evaluación que PISA realiza en otras materias. Dicho marco se caracteriza por tratar de evaluar conocimientos y competencias de historia aplicados a situaciones y problemas de la vida real. La prueba es una variante mejorada de la que figura en el anexo web³ del libro *Pensamiento histórico y evaluación de competencias* (Domínguez, 2015), inicialmente elaborada con el objetivo de ser un ejemplo práctico de la tesis central allí defendida: que lo que denominamos "pensar históricamente" (*Historical thinking*) conlleva la utilización de conceptos metodológicos o "de segundo orden" (fuente y prueba, causas, interpretaciones históricas, etc.), que podemos considerar "competencias históricas", similares a los procesos o competencias científicas, en las que PISA basa sus pruebas de esta materia (Domínguez, 2015:22-25).

En el siguiente apartado se explica cómo se han desarrollado y concretado en destrezas cognitivas evaluables cada una de las competencias y conceptos metodológicos que forman el pensamiento histórico. Para ilustrarlo, se incluyen como ejemplo las destrezas que desarrollan el concepto de explicación contextualizada o empatía histórica con algunos ejemplos de preguntas para su evaluación

En el apartado tercero se explican los rasgos básicos de la prueba: dos grandes unidades o bloques de evaluación, “Emigrantes españoles de ayer y de hoy” (en adelante, EM) y “El trabajo infantil durante la Revolución Industrial y en la actualidad” (en adelante, TI); la presencia de materiales de estímulo sobre los que ha de trabajar el estudiante; los tipos de preguntas o ítems (abiertas, de opción múltiple, etc.); y las competencias y destrezas que cada pregunta pretende evaluar.

En el cuarto apartado se describen las distintas fases del estudio piloto: desde su elaboración hasta el trabajo de campo final en el que han participado unos 800 estudiantes de entre 16 y 18 años. A lo largo de estas etapas, se detalla la validación cualitativa de la prueba por expertos, la revisión tras una breve prueba pre-piloto, y las dos fases que ha tenido el estudio de campo.

En el quinto y último apartado se presentan los resultados provisionales de la primera fase del estudio piloto, analizados estadísticamente, de acuerdo con la Teoría de Respuesta al Ítem (en adelante, TRI). Los resultados permiten concluir provisionalmente que la prueba es válida desde un punto de vista técnico al tener un coeficiente de fiabilidad alto y presentar una distribución de ítems, con niveles variados de dificultad.

1. El contexto teórico de la prueba: un marco para evaluar el pensamiento histórico al final de la ESO.

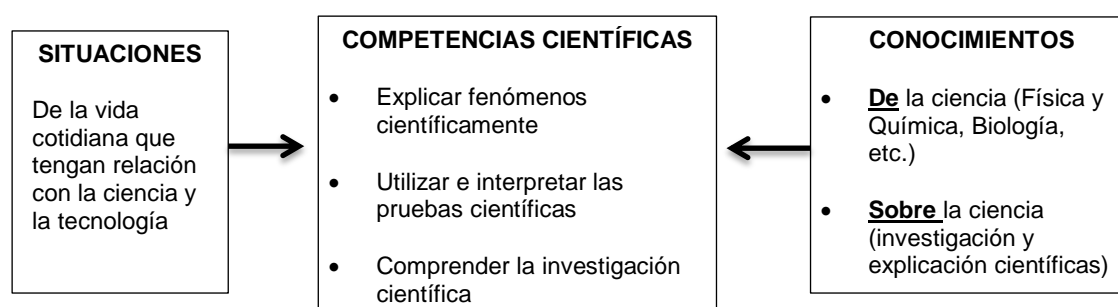
El marco teórico para la evaluación que fundamenta nuestra prueba asume como punto de partida que lo que se denomina “pensar históricamente” o “pensamiento histórico”, se adapta muy bien al marco competencial que utiliza PISA para sus pruebas y, en particular, el que utiliza para evaluar las ciencias experimentales (OCDE, 2013).

1.1. Componentes del marco para la evaluación en PISA

PISA no evalúa conocimientos y destrezas comunes a los currículos nacionales, sino “destrezas que se consideran esenciales para la vida futura”. De ahí su insistencia en evaluar no ciencias sino competencia científica, “para destacar la importancia que la evaluación PISA de ciencias pone sobre la aplicación del conocimiento científico en el contexto de la vida diaria, por comparación con la simple reproducción de conocimiento científico escolar tradicional” (OCDE, 2013:128). Para ello el marco de evaluación que PISA utiliza en ciencias distingue tres componentes (Figura 1).

Las “situaciones” son los hechos concretos sobre los que cada ítem o pregunta trata de comprobar si el alumno sabe aplicar lo aprendido y no solo reproducirlo; los “conocimientos” incluyen las áreas de contenidos habituales, pero añaden además una reflexión meta-científica sobre cómo es y en qué se basa dicho conocimiento; pero el tercer componente, las “competencias”, es la dimensión central de la evaluación: son los procesos u operaciones cognitivas que el alumno debe poner en juego para responder o resolver el problema que le plantea cada pregunta. PISA distingue en ciencias tres grandes competencias o procesos: “Explicar fenómenos científicamente”, es decir, aplicar conocimiento de la ciencia para explicar, predecir, interpretar un fenómeno concreto (por ejemplo, es lavar a alta temperatura la ropa de hospital para reducir el contagio de enfermedades porque así mueren muchas bacterias); “utilizar las pruebas científicas” equivales a saber fundamentar en la información o los datos empíricos disponibles una conclusión o argumento científico; por último, “comprender la investigación científica” significa por ejemplo, distinguir lo que es posible investigar científicamente de lo que no, o reconocer la variable que hay que aislar, en un experimento determinado, etc.

Figura 1. Componentes de la evaluación de las ciencias experimentales en PISA.



Domínguez, 2015:259

¿Es posible adoptar este modelo para diseñar una prueba de evaluación de la competencia histórica? Creemos que sí, si tomamos por tal competencia lo que hoy día solemos entender por “pensar históricamente” (*historical thinking*).

1.2. Historical thinking o pensamiento histórico

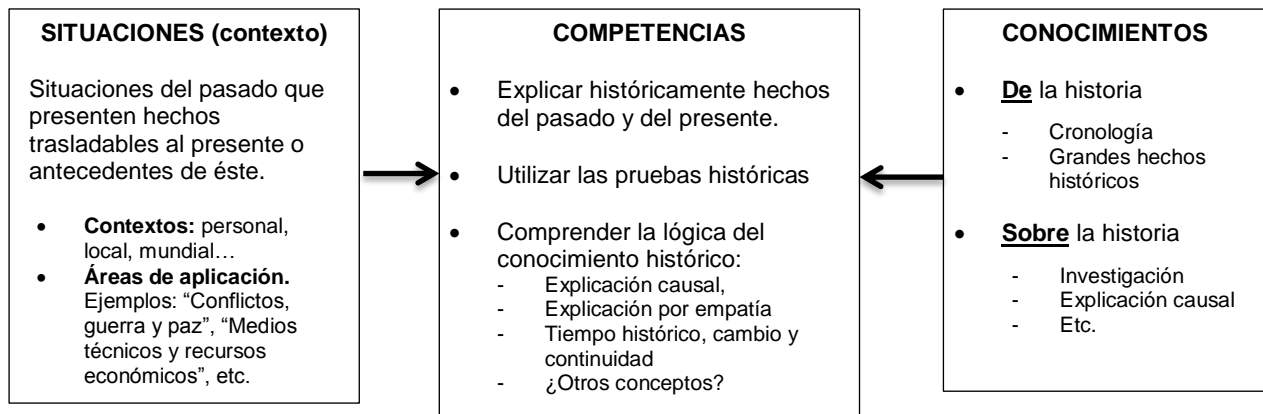
La expresión “pensamiento histórico” (*Historical thinking*) parece haberse convertido en una especie de paradigma de referencia en la didáctica de la historia. Con ella se viene a denominar un aprendizaje de la disciplina que junto a los contenidos sustantivos habituales (saber lo ocurrido en el pasado) otorga una gran importancia a entender cómo sabemos que eso ocurrió (los métodos y formas de trabajo de los historiadores). Una mayoría de autores están de acuerdo en que tales métodos y formas de trabajo de la historia se concretan en un listado básico de “conceptos de segundo orden” o metodológicos como “fuente y prueba”, “causas y consecuencias”, “empatía histórica”, “cambio y continuidad”, entre otros. En general, se puede decir que la gran mayoría de los autores coinciden en estos cuatro conceptos, pero discrepan al incorporar otros conceptos. Así, los autores británicos, entre los que destacamos por su visión de conjunto los trabajos de Lee (2005) y Lee y Shemilt, 2003 (entre otros) añaden a esos cuatro el de *accounts*, que podemos traducir como “relatos e interpretaciones”. Seixas, posiblemente el autor que más y mejor ha popularizado la expresión *historical thinking*, añade a los cuatro conceptos anteriores los de “relevancia” y “dimensión ética” (Seixas y Morton, 2013). Otros autores, y en especial los del área germánica, dan especial atención al concepto de “conciencia histórica” (Rüsen, 2004; Körber y Meyer-Hammer, 2015; Eliasson y otros, 2015).

Entre los distintos conceptos mencionados, nuestro marco de evaluación (Domínguez, 2015:62-63), incluye sólo los cuatro más aceptados: “fuentes y pruebas”, “causas y consecuencias”, “empatía histórica”, y “cambio y continuidad”. Sin embargo no todos tienen igual importancia en la prueba. Duquette ha hecho una distinción entre ellos que coincide en parte con la que PISA hace en su marco para la evaluación de la ciencia (Figura 1): “Pensar históricamente (...) incluye dos elementos principales: perspectiva histórica y método histórico (Laville, 2004)” (Duquette, 2015:52-53). Esta autora incluye dentro de *perspectiva histórica* los conceptos de “relevancia”, “continuidad y cambio”, “causas y consecuencias”, y “empatía histórica” entre otros (los que nosotros situamos dentro de “comprender la lógica de la historia”), mientras que la “utilización de las fuentes como pruebas históricas” constituye el núcleo del método histórico. Por esta razón, entre otras, “fuente y prueba” tiene mayor importancia y peso en nuestro marco de evaluación de la historia (Figura 2 y Tabla 2).

1.3. Un posible marco para evaluar la historia

Siguiendo el modelo de PISA, el marco que proponemos aparece sintetizado a continuación.

Figura 2. Componentes de la evaluación del pensamiento histórico.



Domínguez, 2015:36

No es fácil encontrar en historia *situaciones* aplicables a la vida diaria, como ocurre en las ciencias experimentales. Por eso, en nuestra prueba estas se han creado a partir de la comparación entre situaciones del pasado y del presente: por ejemplo la unidad EM se centra en la emigración española a Europa en los años 1960 y en la actualidad, mientras que la TI analiza la explotación infantil durante la Revolución Industrial en Gran Bretaña y la compara con situaciones del presente.

Con respecto a los *conocimientos*, una prueba competencial como la que aquí se propone no puede, ni mucho menos, cubrir todos los temas de historia contemporánea incluidos en el currículo de la asignatura en 4º de ESO. Como hace PISA, nuestra prueba ha optado por escoger situaciones y problemas para cuya solución sea necesario un conocimiento comprensivo mínimo de algunas épocas y hechos destacados de la historia contemporánea. Así, por ejemplo, para EM: la cronología de las grandes etapas en España desde la Guerra Civil hasta hoy, la postguerra y el desarrollismo en los años 1960, las transformaciones sociales, políticas y culturales de España desde esa década hasta la actualidad y, finalmente, el efecto que la integración en la UE ha tenido sobre nuestra economía desde los años 1990 hasta la crisis presente. Aunque en la prueba algunas preguntas evalúan el recuerdo de ciertos conocimientos históricos, la gran mayoría evalúan la aplicación de tales conocimientos a través de las competencias, a las que nos referimos ahora.

Con respecto a las tres grandes *competencias históricas* (Figura 2), *explicar históricamente...* es la más directamente orientada a evaluar esa capacidad de aplicar conocimientos históricos (contenidos) a los hechos o problemas que hay que explicar o resolver. Así, en EM, como se ha explicado, es imprescindible un mínimo conocimiento de la guerra civil y del periodo de la Dictadura de Franco para situar y comprender mejor la emigración a otros países europeos de muchos españoles en los años 1960. Por su parte, la competencia *utilizar las pruebas históricas* engloba las destrezas que constituyen el método historiográfico basado en las fuentes, sobre todo primarias, y al que solemos referirnos simplemente con el concepto metodológico de "Fuente y prueba". Incluyen dos grandes tipos de destrezas cognitivas estrechamente unidas: destrezas de comprensión lectora y comunicación (obtener información explícita o por inferencia, interpretar y evaluar la información, sintetizar y exponer...) y destrezas historiográficas (análisis crítico, contextualizar, y cotejar y evaluar las fuentes, en tanto que pruebas o evidencias sobre las que basar lo que se afirma del pasado). Finalmente, la competencia *comprender la lógica del conocimiento histórico*, calificada como "perspectiva histórica" por Duquette, incluye numerosas destrezas relacionadas con conceptos de segundo orden como "Explicación causal", "Empatía histórica" y "Tiempo, cambio y continuidad". A estos conceptos cabría añadir otros como "Relatos e

interpretaciones”, “Dimensión ética”, etc., pero pensando en estudiantes de secundaria obligatoria, hemos preferido limitarnos a los tres conceptos más aceptados.

2. Destrezas cognitivas para evaluar cada competencia o concepto metodológico de la historia: el ejemplo de la "empatía histórica"

A la hora de evaluar los anteriores conceptos metodológicos o de segundo orden de la historia se hace necesario distinguir una serie de destrezas cognitivas en cada uno. A modo de ilustración, en este artículo nos referiremos brevemente a las correspondientes al concepto de explicación histórica contextualizada o por empatía (*historical empathy*)⁴ que es, junto con la explicación causal, el otro gran instrumento de explicación histórica. Las causas satisfacen una gran parte de la explicación de por qué ocurren los hechos, pero no la completan, porque dejan sin explicar muchos de los comportamientos, decisiones, acciones o reacciones de los seres humanos que intervinieron en esos hechos. Para explicar esto necesitamos comprender -no necesariamente compartir- las ideas, creencias, expectativas, etc. de esas personas o agentes en ese momento del pasado. A esto denominamos "empatía histórica" o explicación histórica contextualizada (Domínguez, 2015:139 y ss.).

Como hace con los otros conceptos metodológicos, el libro de Domínguez dedica un capítulo a analizar en profundidad la explicación histórica contextualizada. Se trata de un triple análisis que incluye el punto de vista de la filosofía y la metodología de la historia (Dilthey, Collingwood, Schaff), el punto de vista historiográfico, es decir, cómo utilizan algunos historiadores la empatía para explicar ciertos hechos del pasado y, muy en especial, desde el punto de vista de la didáctica de la historia, donde se pone especial atención en las investigaciones sobre su aprendizaje, estrategias de enseñanza y propuestas para evaluar su aprendizaje (Lee, Shemilt, Wineburg, Seixas, entre otros). Sobre la base de ese triple y pormenorizado análisis, Domínguez propone un marco de destrezas cognitivas, directamente evaluables, que desarrollan y concretan el concepto global de empatía histórica, proporcionando así un mapa detallado para el diseño y formulación de ítems que evalúen dicho concepto (ver Tabla 1).

Tabla 1. Destrezas cognitivas en la explicación contextualizada o por empatía histórica.

Identificar y analizar desde el presente (desde la perspectiva de un historiador en el presente) las circunstancias de un hecho del pasado (decisión, acción, comportamiento, actitud...).

Identificar y explicar desde la perspectiva de su época (ideas, mentalidad, creencias, escala de valores, etc. de la gente en aquel momento) un hecho del pasado (decisión, acción, comportamiento, actitud...).

Analizar e interpretar la relación entre las ideas de la gente del pasado (sus intenciones, mentalidad, actitudes, etc.) y las formas de vida de su época (condiciones materiales, organización social y política, y creencias).

Distinguir y comparar las ideas de hoy y de otra época (mentalidad, escala de valores, etc.) en relación con determinados hechos o cuestiones.

Adaptado de Domínguez (2015:164), Cuadro 47.

Ejemplos de preguntas para evaluar la explicación contextualizada o empatía histórica

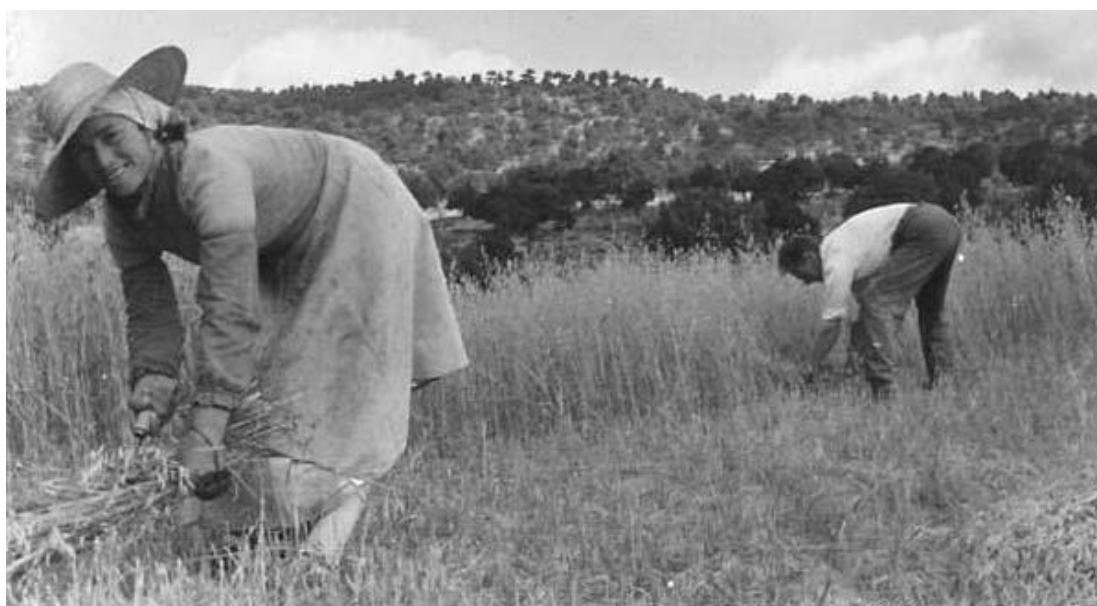
Conviene señalar de partida que a la hora de formular preguntas que permitan evaluar la explicación contextualizada o por empatía, las más idóneas no son preguntas de tipo descriptivo (*imagina que eres uno de los indígenas americanos que vio desembarcar a Colón, cuenta tus impresiones a...*) sino aquellas que planteen al estudiante un problema bien definido, cuya solución

requiera esencialmente adoptar la perspectiva del agente histórico, algo sobre lo que la información del estímulo debe incluir suficientes claves de respuesta. A este respecto también es oportuno insistir en que es más adecuado que el alumno no conozca la solución o conclusión final del problema, por ejemplo, la decisión, actitud, etc. que finalmente asuman los agentes en cuestión.

Veamos a continuación ejemplos de preguntas de la unidad “Emigrantes españoles de ayer y de hoy (1950-2014)” de nuestra prueba, que tratan de evaluar las destrezas propuestas en la Tabla 1. Esta unidad de evaluación se basa en cinco fuentes o documentos de estímulo que presentan información ficticia sobre la emigración a Francia en 1960 de una familia española de origen rural, así como otros datos reales sobre los movimientos migratorios en España tras la crisis de 2008.

Para evaluar la destreza “Identificar y analizar desde el presente (...) las circunstancias...” (Tabla 1) se utilizan preguntas que tratan de evaluar si se comprenden las circunstancias o situación de la gente desde la perspectiva del historiador de hoy. Así, por ejemplo, a partir del estímulo de la Figura 1, se le pregunta al estudiante de qué manera la foto *le ayuda a comprender los motivos de esas personas para emigrar de España*. De este modo se vinculan claramente las circunstancias de la gente de la época con su decisión o no de emigrar.

Figura 3. Siega con hoz (provincia de Teruel, hacia 1950).



<http://www.20minutos.es/museo-virtual/foto/8355/?contexto=temas&tema=agricultura>

La destreza “explicar desde la perspectiva de su época (ideas, mentalidad, etc.) un hecho del pasado (decisión, actitud, etc.)” es sin duda la destreza más importante para evaluar la explicación contextualizada. En el caso de nuestra unidad *Emigrantes* la pregunta clave fue:

*Pregunta 6: Supón que estamos en 1959, meses antes de la partida del abuelo Carmelo a Francia. No tienen hijos y hablan de si es mejor emigrar juntos los dos o sólo él. Trata de ponerte en el lugar de Paquita, con la mentalidad que tendrían las mujeres de aquella época. ¿Cuál de estas frases reflejaría **mejor la postura mayoritaria** en la época? **Escoge solo una***

- a. Aquí no hay futuro, en Francia vas a ganar mucho dinero y podrás comprarte el coche que te gusta. Yo me quedo aquí y así ahorraremos más.*
- b. Yo me quiero ir contigo, lo más importante es estar juntos. Mi madre tiene aquí a mi hermano Mingo que cuidará de ella.*
- c. Es muy duro separarnos, pero tú tienes que traer dinero a casa y yo no puedo dejar sola a mi madre, que se ha hecho muy mayor.*
- d. Me voy contigo, en Francia podremos rehacer nuestra vida, pues desde la muerte del niño no pienso en otra cosa que en volver a tener hijos.*

En esta pregunta se requiere interpretar y contextualizar la decisión de la abuela Paquita desde la perspectiva y mentalidad de las mujeres en aquellos años. En particular, su atención preferente a los familiares necesitados, que impedirían a Paquita abandonar a su madre anciana, aunque esta quedara en compañía de su otro hijo Domingo. Como otras preguntas de opción múltiple, ésta va complementada de una pregunta abierta (explica la opción elegida, considerando la mentalidad y el papel de la mujeres en la época), que tiene por objeto confirmar o no y matizar la respuesta del estudiante.

La unidad “Emigrantes” incluye también preguntas para evaluar otras destrezas relacionadas con este tipo de explicación contextualizada, como muestra la Tabla 2 que recoge las destrezas que tratan de evaluar todas las preguntas de la unidad, pero con lo descrito hasta aquí creemos que es suficiente para que el lector entienda de qué forma cada competencia y concepto de pensamiento histórico como “utilizar pruebas históricas”, “explicación contextualizada”, “tiempo y cambio”, etc. pueden evaluarse a través de destrezas cognitivas específicas.

3. Estructura y características de la prueba de evaluación

La elaboración de nuestro marco teórico para la evaluación del pensamiento histórico se hizo a la par que se diseñaba la prueba de evaluación, de modo que ambas tareas se complementaron de manera provechosa. La prueba fue pensada para estudiantes de final de la Educación Secundaria Obligatoria, que en España tienen edades comprendidas entre los 16 y 18 años, lo que delimitaba tanto los temas o contenidos (época contemporánea) como el nivel de dificultad de estos. La forman dos grandes unidades o bloques de evaluación: “Emigrantes españoles de ayer y de hoy (1950-2014)” (EM) y “El trabajo infantil en la Revolución Industrial y en la actualidad (1800-2014)” (TI). Ambas constan de dos cuadernillos, uno con los documentos de estímulo y el otro con el cuestionario de ítems y los espacios para la respuesta. Los documentos de estímulo (cinco en EM y nueve en TI) son en su mayoría fuentes primarias que informan sobre hechos concretos: emigración de una familia rural española en 1960, datos sobre movimientos migratorios actuales en España; trabajo de niños en fábricas de hilado durante la Revolución Industrial británica, mapa mundial sobre el trabajo infantil en 2006, etc. Son de distinto formato (escritos, imágenes, gráficos y mapas) y constituyen la base de información que el estudiante deberá utilizar (junto a sus conocimientos) para contestar a la gran mayoría de las preguntas. En cada unidad de evaluación, se abordan hechos del pasado y otros de la actualidad, lo que facilita la aplicación del conocimiento histórico y posibilita la reflexión pasado-presente. Aunque PISA suele utilizar unidades más cortas (entre 1 y 3 estímulos en la mayoría de ellas), pensamos que en historia es preferible utilizar unidades más largas, por varios motivos: rentabilizar el tiempo dedicado a la lectura de los documentos con preguntas que evalúen distintas competencias, disponer de varias fuentes para poder cotejarlas y valorar la información que proporcionan y, finalmente, para contar con información sobre un mismo tema en distintos momentos históricos y poder evaluar el cambio y la continuidad, e incluso la reflexión pasado-presente-futuro.

Para obtener significatividad en los parámetros estimados mediante TRI se necesitaba un número alto de preguntas que permitieran medir varias competencias y distintos niveles de dificultad, dentro de ellas. Así se llegó a un total de 39 preguntas o ítems (22 en EM y 17 en TI) que, aunque se pueden responder en unos 75-90 minutos, por razones prácticas se respondieron en dos periodos de clase de 45 minutos aproximadamente. Las preguntas son de tres tipos: opción múltiple, semi-construidas, y abiertas aunque de respuesta relativamente breve.

En la Tabla 2 se recogen los ítems de la unidad *Emigrantes*, con indicación del tipo de pregunta y la competencia, concepto y destreza que cada una trata de medir. En la última fila de la tabla se recoge en porcentajes la distribución de ítems por tipo de pregunta y por competencia evaluada. En cada unidad de evaluación se han incluido ítems sobre las tres competencias, pero la explicación por empatía sólo se evalúa en esta unidad, mientras que la explicación causal se evalúa en la del “Trabajo Infantil”.

Tabla 2. Tipo de preguntas y competencias evaluadas en la unidad Emigrantes españoles de ayer y de hoy (1950-2014).

Número de pregunta	Tipo de pregunta			Competencias y destrezas evaluadas			
	Opción Múltiple	Abierta	Semi estructuradas	UTILIZAR PRUEBAS HISTÓRICAS	COMPRENDER LA LÓGICA DEL CONOCIMIENTO HISTÓRICO		EXPLICAR HISTÓRICAMENTE
					Explicación contextualizada o por empatía	Tiempo histórico, cambio y continuidad	
1			*			(*) Cronología	*
2		*		* Interpretar+Cotejar			
3.1	*			* Inferir+Analizar			
3.2		*		* Inferir +Analizar			
4		*		(*) Interpretar+Context.			*
5.1	*			* Inferir+ Contextualizar	(*) Adoptar persp. historiador		(*)
5.2			*	(*) Inferir+ Contextualizar	(*) Adoptar persp. historiador		*
6.1	*				* Adoptar persp. pasado		(*)
6.2		*			* Adoptar persp. pasado		(*)
7.1	*						*
7.2	*						*
8.1	*			* Interpretar+Analizar			
8.2			*	* Interpretar+Analizar			
9.1.		*				* Procesos de cambio y continuidad	
9.2	*						*
10	*			* Evaluar+Cotejar			
11.1			*	(*) Interpretar+Cotejar		* Contenido del cambio	(*)
11.2		*		(*) Comunicar +Cotejar		* Contenido del cambio	(*)
12.1	*				* Comparar perspectivas		

12.2		*			(*) Comparar perspectivas	* Contenido del cambio	(*)
13	*				(*) Relacionar perspectivas y formas de vida		*
14		*				(*) Contenido del cambio	*
22	10	8	4	7+ (4)	3 +(3)	4 +(2)	8 + (6)
100%	45,5%	36,3%	18,2%	31,8%	31,8%		36,3%

(*): Competencias evaluadas complementariamente o en segundo lugar.

4. Aplicación del estudio piloto

La elaboración definitiva de la prueba, entre enero de 2014 y marzo de 2015, conllevó una validación cualitativa realizada por expertos en distintas especialidades y otra más cuantitativa y experimental, mediante un limitado trabajo de campo realizado con 49 alumnos de 4º de ESO. Ambas permitieron detectar problemas en el diseño y formulación de las preguntas, lo que llevó a hacer una importante revisión final de la prueba, con numerosos cambios que afectaron a la reformulación, replanteamiento, e incluso al cambio y sustitución de algunos ítems⁵. Por otra parte, este estudio pre-piloto permitió comprobar que la gran mayoría de estudiantes no lograban terminar la prueba en un solo periodo de clase de 50 minutos, por lo que se optó por pasar cada una de las unidades de evaluación en distintos periodos lectivos.

El estudio de campo ha tenido como objetivo comprobar la validez y fiabilidad de la prueba en sí misma, como instrumento para evaluar el pensamiento histórico del alumnado. Es decir, el objetivo ha sido evaluar la prueba, no a los estudiantes. Por ello, algunas de las características de este estudio han venido también impuestas por el posterior análisis estadístico TRI: aparte del alto número de preguntas mencionado, se ha aplicado a una muestra de conveniencia, no representativa de ninguna población de estudiantes (ciudad, comunidad autónoma, etc.), aunque razonablemente variada, por sexo y nivel socio-cultural y, sobre todo, numerosa (unos 500 estudiantes), para poder disponer de una base estadística suficiente.

La aplicación de la prueba se ha llevado a cabo en dos fases: la primera en mayo de 2015 y la segunda entre marzo y abril de 2016. En 2015 participaron un total de 316 estudiantes de 4º de ESO pertenecientes a 15 grupos de clase de 10 centros educativos, en 8 municipios de la Región de Murcia con diferentes características geográficas y socioeconómicas. 4 aplicadores pasaron el grueso de las pruebas (13 grupos), que se completaron con la ayuda de otros 3 profesores. Al aplicarse en mayo, los centros facilitaron solo una hora de clase, por lo que cada alumno pudo responder solo a una unidad de evaluación. Por circunstancias diversas, la distribución de las unidades no pudo hacerse de manera equilibrada, resultando que, de los 316 cuestionarios respondidos, 201 (el 64 %) correspondían a EM y 115 (el 36 %) a TI. Además de cuantitativo este desequilibrio fue también cualitativo, pues en la corrección se puede apreciar que un número significativo de los cuadernillos de TI no estaban terminados, lo que podía deberse a que al menos 3 de los grupos de alumnos que respondieron a esta unidad eran grupos de diversificación, en principio por tanto de menor rendimiento, lo que pudo afectar a los niveles de respuesta en esta unidad.

Para las respuestas a las preguntas abiertas y semi-construidas, se elaboró una guía con los criterios de corrección y codificación o puntuación, que fue revisada y consensuada por los cuatro correctores participantes. Los cuadernillos fueron corregidos y codificados por estos cuatro profesores que además trabajaron juntos en todo momento, de manera que las dudas que pudieran

generar algunas respuestas eran discutidas y resueltas de inmediato. Como mostraría después el análisis estadístico, la codificación se realizó con una gran coherencia y fiabilidad y apenas sin errores.

Esta primera fase se ha completado con una segunda aplicación, entre marzo y mayo de 2016, en la que han participado en torno a 500 estudiantes de 4º ESO. Esta vez se han empleado dos periodos de clase, de manera que cada estudiante ha respondido a las dos unidades que forman la prueba completa. Antes de finalizar 2006 esperamos completar la corrección y codificación de los cuadernillos de esta segunda fase del estudio de campo, lo que nos permitirá hacer un análisis TRI más completo de la prueba.

5. Análisis T.R.I. de los resultados obtenidos en la primera fase del estudio piloto

Las respuestas codificadas de los cuadernillos respondidos en la primera fase del estudio de campo han sido analizadas estadísticamente, de acuerdo con la Teoría de Respuesta al ítem (TRI), mediante la aplicación informática “Conquest”.

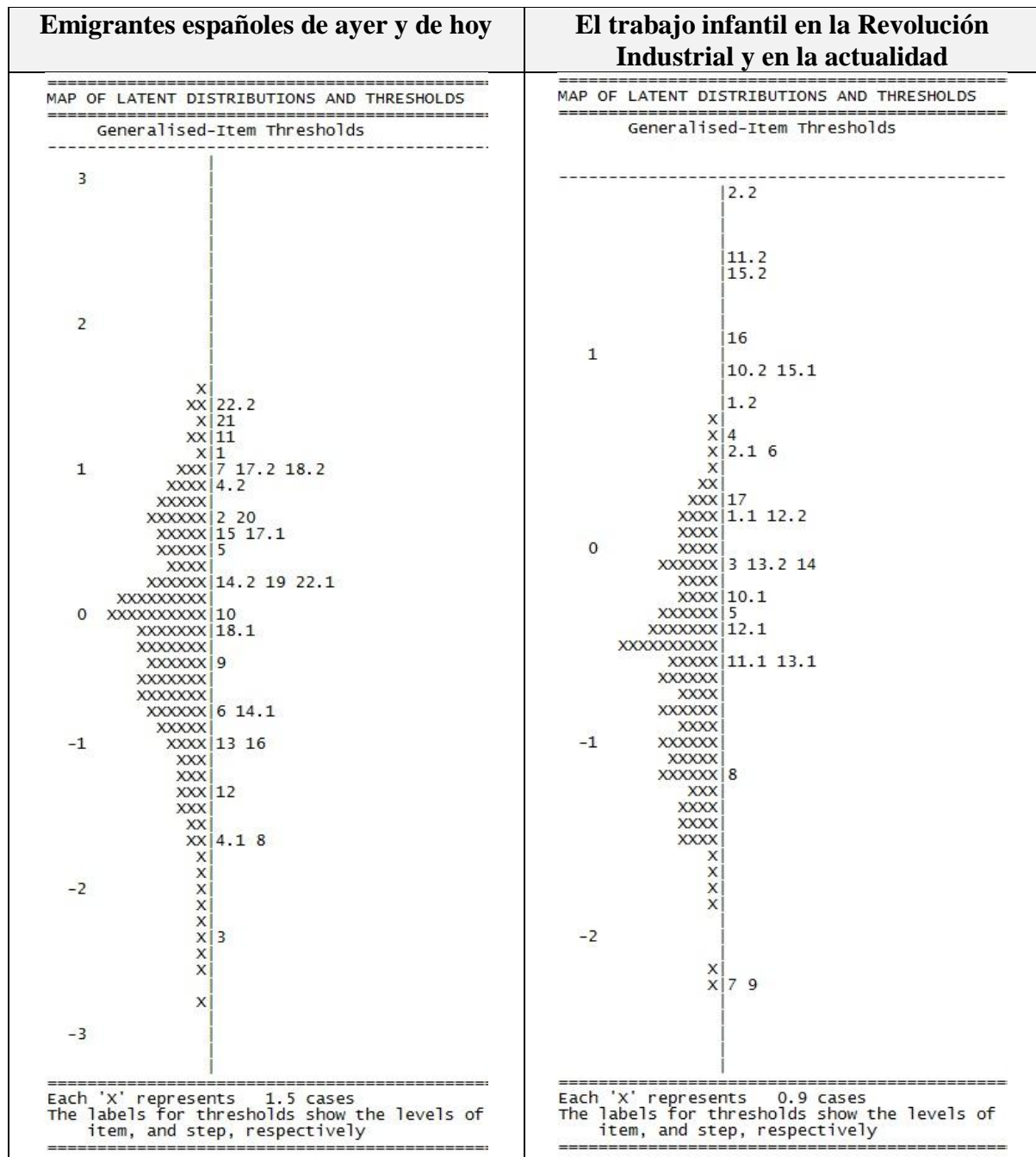
El análisis TRI estima a la vez el grado de dificultad de los ítems y la habilidad de los estudiantes, y esto se hace recíprocamente, es decir, el nivel de dificultad de cada pregunta se obtiene del porcentaje de estudiantes que la responden correctamente (cuantos más lo hagan, más sencilla será la pregunta, y cuantos menos, más difícil) y, de modo similar, la habilidad o rendimiento de cada estudiante se establece en función de las preguntas de nivel más alto a las que responde correctamente de forma consistente (según esto, un estudiante que se encontrara en un determinado nivel habría respondido acertadamente a la mayoría de ítems de menor o igual dificultad, y sólo a unos pocos de mayor dificultad) (Embretson y Reise, 2000)

De los resultados obtenidos de ese análisis pueden destacarse tres aspectos: Grado de discriminación de cada ítem, Coeficiente de fiabilidad, y distribución equilibrada. Con respecto a la **discriminación** (si la pregunta diferencia con coherencia a los estudiantes según su capacidad en el conjunto de la prueba), en EM todos los ítems discriminan bien, mientras que en TI, se detectaron dos preguntas de opción múltiple, que afortunadamente pudieron cambiarse a tiempo para la segunda aplicación de la prueba.

El **coeficiente de fiabilidad** de una prueba permite decir si, en conjunto, todos los ítems de la misma van o no en la misma dirección. En nuestra prueba, la unidad EM tiene un coeficiente alto, del 78%, mientras que en TI, es del 65%, claramente inferior aunque válido.

Finalmente, la **distribución** de los ítems permite mostrar en una misma escala el grado de dificultad de los ítems y las puntuaciones de los estudiantes. En nuestra prueba las puntuaciones de los alumnos siguen una distribución normal que es lo esperado en estos casos. La Figura 2 muestra dos diagramas, uno para EM y otro para TI. El eje vertical marca una escala de dificultad que va de -3 a 3 (mínima y máxima dificultad). A la derecha del eje se ordenan los ítems (indicados con su número correspondiente) y a la izquierda las respuestas obtenidas (cada “x” representa 1,5 casos de respuesta). El diagrama muestra que EM tiene una distribución muy equilibrada de ítems y respuestas, mientras que en TI la mayoría de ítems se sitúan por encima de 0, y pocos alumnos han sido capaces de responderlas. De ello parece deducirse que TI es una unidad más difícil en conjunto que la otra, aunque esa mayor dificultad puede ser aparente y debida en parte a que los estudiantes que la han respondido tengan, de promedio, un nivel de habilidad inferior que los que respondieron a EM. Esto último es probable ya que, como se ha dicho, muchos de los estudiantes pertenecían a grupos de diversificación y dejaron bastantes preguntas sin responder.

Figura 4. Mapas de distribuciones de los ítems y respuestas en cada unidad según niveles de dificultad.



Conclusiones

A juzgar por los primeros resultados analizados, la prueba parece un instrumento apropiado para evaluar las competencias de pensamiento histórico:

- 37 de un total de 39 ítems *discriminan* bien la habilidad de los estudiantes en consonancia con la puntuación que logran en el conjunto de la prueba.
- El *coeficiente de fiabilidad* es alto en la unidad de EM (0,78 sobre 1) y suficiente en la de TI (del 0,65)
- En conjunto, el *mapa de distribución de ítems y puntuaciones* de las dos unidades, permite confiar en que, tras la segunda fase de la aplicación que proporcionará respuestas de todos los estudiantes a las dos unidades, se pueda establecer una escala conjunta de niveles de logro en el pensamiento histórico, que distingan las diferentes destrezas cognitivas y competencias. De ese

modo, este tipo de prueba sería no solo un instrumento válido para medir el pensamiento histórico de los estudiantes, sino que serviría también como modelo de ejercicios para investigar los niveles de aprendizaje en la adquisición de dicho pensamiento y, por supuesto, como modelo de actividades didácticas para su enseñanza.

Bibliografía

- Domínguez, J. (2015).** *Pensamiento histórico y evaluación de competencias*. Barcelona: Graó.
- Domínguez, I.; Arias, L.; Sánchez, R.; Egea, A. & F. J. García (2016).** Cómo evaluar el pensamiento histórico en la ESO: primeros resultados de una prueba piloto. En López Facal, R. (ed.). *Ciencias sociales, educación y futuro. Investigaciones en didáctica de las ciencias sociales. Libro de Actas del Simposio (176-187)*. Santiago de Compostela: Red I4-USC.
- Duquette, C. (2015).** Relating Historical Consciousness to Historical Thinking through Assessment. In Ercikan, K. & P. Seixas (Eds.). *New Directions in Assessing Historical Thinking* (51-63). Londres-New York: Routledge.
- Eliasson, P.; Alvén, F.; Yngvéus, C. A. & D. Rosenhud (2015).** Historical Consciousness and Historical Thinking Reflected in Large-Scale Assessment in Sweeden. In Ercikan, K. & P. Seixas (Eds.). *New directions in Assessing Historical Thinking* (171-182). New York, NY: Routledge.
- Embretson, S. E. & S. P. Reise (2000).** *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Körber, A. & J. Meyer-Hammer (2015).** Historical Thinking, Competencies and their Measurement: Challenges and Approaches. In Ercikan, K. & P. Seixas (Eds.). *New directions in Assessing Historical Thinking* (89-101). Londres-New York: Routledge.
- Laville, C. (2014).** Historical Consciousness and Historical Education: What to Expect from the First for the Second. In Seixas, P. (Ed.). *Theorising Historical Consciousness* (165-182). Toronto: University of Toronto Press.
- Lee, P. J. (2005).** Putting principles into practice: Understanding history. In Donovan, M. S. & J. D. Bransford (eds). *How Students Learn: History in the Classroom*. US National Research Council, Washington DC: National Academies Press.
- Lee, P. J. & D. Shemilt (2003).** A scaffold, not a cage: Progression and Progression Models in History. *Teaching History*, 113, 13-23.
- OECD (2013).** *PISA 2012: Assessment and Analytical Framework. Mathematics, Reading, Science, Problem Solving and Financial Literacy*. OECD Publishing. Recuperado de: https://www.oecd.org/pisa/pisaproducts/PISA%202012%20framework%20De-book_final.pdf
- Rüsen, J. (2004).** Historical Consciousness: Narrative Structure, Moral Function and Ontogenetic Development. In Seixas, P. (Ed.). *Theorizing Historical Consciousness* (63-85). Toronto: University of Toronto Press.
- Seixas, P. & T. Morton (2013).** *The Big Six Historical Thinking Concepts*. Toronto: Nelson.

Notas

¹ Para la elaboración de este artículo hemos utilizado ampliamente el texto de nuestra comunicación “Cómo evaluar el pensamiento histórico en la ESO: primeros resultados de una prueba piloto” (ver referencias bibliográficas), pero se incluyen también nuevos contenidos, especialmente en el apartado 2 del artículo.

² Este trabajo ha contado con el apoyo de los proyectos “Competencias sociales para una ciudadanía democrática” (EDU2015-65621-C3-1-R), “La evaluación de las competencias básicas en Educación Secundaria Obligatoria desde las ciencias sociales” (EDU2012-37909-C03-03) y “La evaluación de las competencias y el desarrollo de capacidades cognitivas sobre historia en Educación Secundaria Obligatoria” (EDU2015-65621-C3-2-R), subvencionados por el Ministerio de Economía y Competitividad español y fondos FEDER.

³ La prueba es de libre acceso y puede descargarse de http://www.grao.com/dades/BB308/BB308ANEXO_U1.pdf y http://www.grao.com/dades/BB308/BB308ANEXO_U2.pdf

⁴ El término empatía (*historical empathy*) ha generado controversia en el ámbito anglosajón, por sus connotaciones afectivas y de identificación con las ideas y actitudes de los agentes del pasado. En su lugar se han propuesto términos como “tomar (otra) perspectiva” (*perspective taking*) y “pensamiento históricamente contextualizado”. Nosotros utilizamos aquí el de “explicación contextualizada” (Domínguez, 2015:144,150).

⁵ Se partió de la prueba que figura en el anexo web del libro de Domínguez, 2015.