



Revista mexicana de investigación educativa

ISSN: 1405-6666

ISSN: 2594-2271

Consejo Mexicano de Investigación Educativa A.C.

Marchant, Nicole Abricot; Santelices, María Verónica; Zúñiga, Carmen
Gloria; Castañeda, Lucía Valencia; Miranda-Arredondo, Paloma
Evaluar el pensamiento social: desarrollo de un instrumento de evaluación de desempeño en el aula*
Revista mexicana de investigación educativa, vol. 28, núm. 96, 2023, Enero-Marzo, pp. 187-222
Consejo Mexicano de Investigación Educativa A.C.

Disponible en: <https://www.redalyc.org/articulo.oa?id=14074950008>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

UAEH  redalyc.org

Sistema de Información Científica Redalyc
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso
abierto

EVALUAR EL PENSAMIENTO SOCIAL

*Desarrollo de un instrumento de evaluación de desempeño en el aula**

NICOLE ABRICOT MARCHANT / MARÍA VERÓNICA SANTELICES /

CARMEN GLORIA ZÚÑIGA / LUCÍA VALENCIA CASTAÑEDA / PALOMA MIRANDA-ARREDONDO

Resumen:

Para contribuir a la evaluación en ciencias sociales en el aula, este artículo presenta evidencias de validez sobre un instrumento diseñado siguiendo el método *Bear Assessment System* para retroalimentar el pensamiento social. Participaron 175 estudiantes universitarios que escribieron un discurso sobre un problema social, el cual fue corregido con la rúbrica creada para ello. Como evidencia de su validez se reconocen la fiabilidad entre jueces y su consistencia interna, además de observarse cuatro dimensiones en su estructura de acuerdo con la expectativa teórica. No se identificaron relaciones entre el puntaje y el género, y aunque se esperaba una relación positiva con el desempeño escolar, esta no fue apoyada por la evidencia. Las conclusiones destacan la viabilidad de medir el pensamiento social como un constructo multidimensional para retroalimentar el aprendizaje y presentan ideas para mejorar el instrumento.

Abstract:

To contribute to classroom assessment in social science, this article presents evidence of the validity of an instrument based on the Bear Assessment System to provide feedback on social thinking. The participants were 175 university students who wrote a speech about a social problem, which was scored according to a specific rubric. Recognized evidence of the instrument's validity was provided by judges' reliability and the instrument's internal consistency, in addition to the observed four dimensions in its structure in line with theoretical expectations. No relations were identified between scores and gender, and the evidence did not support the expected positive relationship with scholastic performance. The conclusions emphasize the viability of assessing social thinking as a multidimensional construct for providing feedback on learning. Ideas for improving the instrument are presented.

Palabras clave: aprendizaje social; validez de pruebas; evaluación del aprendizaje; ciencias sociales; aprendizaje de la historia.

Keywords: social learning; test validity; evaluation of learning; social science; study of history.

Nicole Abricot Marchant: profesora de la Pontificia Universidad Católica de Chile. Macul, Santiago, Chile.
CE: neabricot@uc.cl / <https://orcid.org/0000-0003-4539-486X>

María Verónica Santelices: profesora de la Pontificia Universidad Católica de Chile. Macul, Santiago, Chile.
CE: vsanteli@uc.cl / <https://orcid.org/0000-0003-4062-0047>

Carmen Gloria Zúñiga: profesora de la Pontificia Universidad Católica de Chile. Macul, Santiago, Chile.
CE: carmen.zuniga@uc.cl / <https://orcid.org/0000-0001-6031-9687>

Lucía Valencia Castañeda: profesora de la Universidad de Santiago de Chile. Estación Central, Santiago, Chile.
<https://orcid.org/0000-0002-5881-1100>

Paloma Miranda-Arredondo: profesora de la Universidad de Santiago de Chile. Estación Central, Santiago, Chile.
<https://orcid.org/0000-0003-3122-7745>

*Este trabajo fue financiado por la Agencia Nacional de Investigación y Desarrollo (ANID) / Programa de Becas / Doctorado Nacional /2018-21181415.

Introducción

Durante las últimas décadas, las necesidades de equidad, democracia y justicia se han instalado como demandas transversales en América Latina. Esto ha tensionado de forma directa el rol de la escuela en la formación de jóvenes capaces de participar activamente de los desafíos sociales (Unesco, 2016). A pesar de que la formación ciudadana corresponde a todas las disciplinas del aprendizaje escolar, las ciencias sociales son un área privilegiada en este propósito (Hahn, 2020; Ross, 2017; Levstik y Tyson, 2008).

Desafortunadamente, las evidencias en la literatura demuestran que la enseñanza de las ciencias sociales, más que favorecer el desarrollo de la formación ciudadana, resulta lejana para las y los estudiantes¹ y basada en la reproducción de contenidos (Monte-Sano y Quince, 2021; Torrez y Claunch-Lebsack, 2014). Existe amplia discordancia entre los discursos de los docentes que declaran promover el aprendizaje crítico y reflexivo y la observación de sus prácticas en el aula, que revela un enfoque centrado en el profesor.

En evaluación, este enfoque se caracteriza por procedimientos centrados en el uso de la memoria y desalineados de los propósitos de aprendizaje, los cuales normalmente están basados en instrumentos de respuesta cerrada que no favorecen el desarrollo de habilidades (Smith, Breakstone y Wineburg, 2018; Torrez y Claunch-Lebsack, 2014). Por ejemplo, un estudio desarrollado con 25% de la población docente de la región de Murcia, España, indica que el 97% del profesorado de Historia y Geografía declaró que la evaluación a través de exámenes es la práctica más extendida y utilizada en el subsector (Alfageme-Gonzalez, Monteagudo y Miralles, 2016). Mientras que Grant (2017) es enfático en señalar que este tipo de evaluaciones a menudo carece de validez al focalizarse en listar personajes, lugares o eventos.

Debido a lo anterior, la literatura destaca la necesidad de estudiar empíricamente evidencias de validez de los instrumentos de evaluación diseñados para retroalimentar el aprendizaje en el aula de ciencias sociales (Smith, 2018; Smith, Breakstone y Wineburg, 2018). Esta investigación propone un instrumento de este tipo, particularmente para la evaluación del pensamiento social.

Evidencias de validez sobre las puntuaciones e instrumentos de evaluación en ciencias sociales

Entendiendo que la validez es un argumento derivado de evidencias que sostienen el uso e interpretación de las puntuaciones de un test (Kane,

2020), es posible recolectar evidencias sobre el contenido, los procesos de respuesta, la estructura interna, la relación con otras variables y las consecuencias de uso de un instrumento (American Educational Research Association, American Psychological Association y National Council on Measurement in Education, 2014). Algunos estudios se han concentrado en analizar este tipo de evidencias para las puntuaciones de instrumentos de evaluación que pretenden medir y retroalimentar algunas de las habilidades del pensamiento histórico o geográfico, que conforman el pensamiento social.

Debido a su extendido uso, los análisis han abordado principalmente instrumentos de respuesta múltiple. Smith (2017) y Reich (2011) examinaron evidencias sobre los procesos de respuesta de test estandarizados nacionales a través de procesos de pensamiento en voz alta, reconociendo que este tipo de instrumentos activaba constructos irrelevantes, tales como rememoración, comprensión de lectura y estrategias de respuesta a los test. A raíz de lo anterior, otras investigaciones analizaron evidencias de validez de instrumentos diseñados particularmente para evaluar el constructo de pensamiento histórico a través de preguntas de opción múltiple. Estos trabajos concluyen que, a diferencia de los test estandarizados nacionales, los de opción múltiple logran evocar habilidades de pensamiento histórico, aunque siguen movilizando constructos irrelevantes en las respuestas de los estudiantes (Smith, 2018; Smith, Breakstone y Wineburg, 2018; Reisman, 2012).

En respuesta a lo anterior, otros estudios han analizado formas alternativas para evaluar el pensamiento histórico, el pensamiento geográfico o algunas de sus habilidades de forma particular. Estos trabajos se han dedicado a recopilar evidencias sobre procedimientos evaluativos basados en desempeño durante la implementación de intervenciones pedagógicas que miden logros del estudiantado utilizando rúbricas (Huijgen, Holthuis, Van Boxtel, Van De Grift *et al.*, 2019; Ercikan y Seixas, 2015; Monte-Sano y De La Paz, 2012).

En esta línea de investigación es posible reconocer las preguntas basadas en documentos, donde se entregan fuentes de información a los estudiantes para que luego elaboren un ensayo para demostrar su dominio del pensamiento histórico. Esta perspectiva ha sido valorada en países como Estados Unidos debido a la autenticidad de su medición, generando incluso test estandarizados (College Board, 2017). Sin embargo, las escasas investigaciones

cuestionan su fiabilidad y su validez (p. ej., índices de consistencia interna entre .4 y .53; y consistencia entre jueces entre .67 y .71; Wainer, 2011; Grant, Gradwell y Cimbricz, 2004). De La Paz y Wissinger (2015) señalan que hay diferencias estadísticamente significativas ($F(1, 101) = 6.206$, $MSE = 12.542$, $p = .01$) en las puntuaciones de estas tareas dependiendo del género textual que solicitan y del nivel de conocimiento histórico de quienes responden; así como de la longitud de los escritos ($F(1, 101) = 5.46$, $MSE = 24147.39$, $p = .02$). De La Paz, Ferretti, Wissinger, Yee *et al.* (2012) reconocen mejores puntuaciones en habilidades de pensamiento histórico en los buenos escritores que en aquellos menos aventajados ($F(3, 87) = 13.59$, $MSE = 4.48$, $p = .000$). Este tipo de evidencias supone desafíos importantes para estos instrumentos estandarizados.

El reciente estudio de Supianto, Kumaidi y Suryono (2021) presenta mejores coeficientes de fiabilidad (.9) y evidencias positivas sobre el contenido y la estructura interna de las puntuaciones de un instrumento de estas características, pero a pesar de que también promueve la autenticidad de las evaluaciones, se limita a evaluar conocimiento histórico.

Los esfuerzos descritos coinciden con el interés por avanzar en una evaluación auténtica del aprendizaje de las ciencias sociales (Pearcy, 2017). Sin embargo, además de las dificultades presentadas, la mayoría se centra en evaluar el pensamiento histórico usando eventos del pasado, sin reconocer que los propósitos de aprendizaje se relacionan más bien con desarrollar habilidades de orden superior para comprender el mundo sociocultural desde las distintas disciplinas (Monte-Sano y Quince, 2021; Percy, 2017). Este artículo propone un instrumento para evaluar ese tipo de aprendizaje, considerando los desafíos ya descritos respecto de la medición del aprendizaje en el aula de ciencias sociales.

¿Qué se debe aprender en el aula de ciencias sociales?

Existen dos perspectivas que han caracterizado los aprendizajes en ciencias sociales. Por una parte, diversos autores han destacado que aprender una ciencia social, como historia o geografía, implica alfabetizarse en ella mediante el uso de un lenguaje académico y una forma de pensamiento disciplinar específico (Monte-Sano y De La Paz, 2012). Esta corriente ha generado una amplia producción teórica y empírica y su producción ha estado centrada en el desarrollo del pensamiento histórico (Henríquez, Carmona, Quinteros y Garrido, 2018; De la Paz, Monte-Sano, Felton,

Croninger *et al.*, 2017; Monte-Sano y De La Paz, 2012). Su propósito principal se centra en promover el pensamiento de una disciplina en particular, más que una integración disciplinar.

En otra perspectiva se encuentran los estudios sociales, que plantean que la enseñanza y el aprendizaje de las ciencias sociales debe abordarse integradamente. Su fin principal es formar ciudadanos capaces de analizar críticamente la complejidad de la realidad social (McGlinn y Mason, 2017; Saye, 2017; Ross, 2017; Gerwin, 2003). Esto se materializa en la propuesta de los Estándares Curriculares para los Estudios Sociales (National Council for the Social Studies, 2013) de Estados Unidos. Allí se definen cuatro dimensiones de estándares relacionadas con el cuestionamiento de la sociedad, el análisis desde conceptos de las disciplinas, la interpretación de fuentes, y la comunicación y toma de decisiones sobre problemáticas sociales. Sin embargo, aunque el documento reconoce habilidades integradoras de las disciplinas, no se identifica un constructo general que permita visibilizar el aprendizaje que los estándares esperan desarrollar.

Desde ese desafío, la corriente iberoamericana ha defendido que el propósito principal de la enseñanza en el área es el desarrollo del *pensamiento social*. Este permite a las personas actuar como ciudadanos críticos de su realidad (Gutiérrez y Pagés, 2018; Murillo y Martínez, 2014; Pagés, 2002). Este pensamiento de orden superior permitiría que los estudiantes conciban la realidad como una síntesis compleja, que contextualicen la información social en sus múltiples dimensiones y que comprendan su realidad desde una perspectiva crítica y participativa (Gutiérrez y Pagés, 2018; Pagés, 2002; Pipkin y Sofía, 2005). Para promover este pensamiento, Pagés y Santisteban (2011) enfatizan en la necesidad de situar al estudiante frente a problemas sociales relevantes y así abordar el aprendizaje de la complejidad social desde lo cotidiano.

Hacia una propuesta interdisciplinaria del aprendizaje en ciencias sociales

Según Gutiérrez y Pagés (2018), el pensamiento social se organiza en cuatro dimensiones denominadas “descripción del conocimiento social (CS)”, “explicación y justificación del CS”, “interpretación del CS” y “argumentación del CS”. Cada una está constituida por la integración de diversas habilidades disciplinares. De esta forma, al describir un problema social se espera que los estudiantes lo contextualicen en términos temporales, espaciales y desde el reconocimiento de agentes y dimensiones; cuando explican un problema,

deberían desarrollar una cadena con diversas causas y consecuencias. Además, se busca que el alumnado interprete críticamente la evidencia disponible y que pueda definir una posición y plantear propuestas de solución frente a las problemáticas (Gutiérrez y Pagés, 2018; Murillo y Martínez, 2014).

Los contenidos descritos en cada una de las dimensiones del pensamiento social derivan, en su mayoría, de investigaciones sobre el pensamiento histórico, y una menor cantidad, de otras disciplinas o se posicionan en la perspectiva de los estudios sociales. Sin embargo, la complejidad que implica la integración de las disciplinas, la segmentación de evidencias y la falta de acuerdos respecto de qué se debe evaluar (Saye, 2017) son desafíos relevantes para la evaluación en el aula. Esta investigación busca reportar evidencias de validez de las puntuaciones de un instrumento diseñado para evaluar el pensamiento social de los estudiantes al egresar del sistema escolar, con el fin de guiar la retroalimentación del aprendizaje y aportar antecedentes que permitan avanzar en la consolidación del constructo en la evaluación de aula.

¿Cómo se describen las dimensiones del pensamiento social?

En la actualidad no existen instrumentos que permitan evaluar el pensamiento social como un constructo en sí mismo. No obstante, existe evidencia segmentada en diversos estudios respecto de las habilidades que lo componen en las diversas disciplinas (Monte-Sano y De La Paz, 2012; Reisman y Wineburg, 2008; Siegfried, Krueger, Collins, Frank *et al.*, 2010).

La evidencia sobre habilidades específicas del pensamiento histórico, geográfico o económico aporta antecedentes sobre las dimensiones que componen el constructo de pensamiento social. A continuación, se describe esa evidencia para cada una de las dimensiones que, hipotéticamente, conforman el constructo de pensamiento social que se investiga en esta propuesta.

La dimensión *contextualización de las problemáticas sociales* (Huijgen, Van Boxtel, Van de Grift y Holthuis, 2017; Huijgen, Van de Grift, Van Boxtel y Holthuis, 2018; Huijgen *et al.*, 2019; Reisman y Wineburg, 2008) representa aquello que hacen las personas cuando describen los problemas sociales y los sitúan en un marco temporal y espacial (Smith, 2014; Reagan, 2008; Miralles, Gómez y Monteagudo, 2012). Algunos autores presentan este proceso como parte de la explicación (Lee, 2009), pero otros lo reconocen como una habilidad en sí misma. Esta permitiría reconocer la complejidad de las interacciones entre los personajes, actores

y entidades sociales (Huijgen *et al.*, 2017 y 2018) en los ámbitos políticos, sociales, económicos y culturales (Ortuño, Ponce y Serrano, 2016; Siegfried *et al.*, 2010).

La *explicación de los problemas sociales* se manifiesta cuando las personas argumentan por qué suceden los problemas y qué consecuencias tienen (Lee, 2009). Esto implica determinar las causas, motivaciones, hechos y consecuencias de los problemas sociales (Kemp, 2011). Las causas a las que se alude son diversas (Stoel, Van Drie y Van Boxtel, 2015; Miralles, Gómez y Monteagudo, 2012) y manifiestan distintos cambios y continuidades relacionadas con la problemática (Seixas, 2017). De igual forma, las personas empatizan, es decir, reconocen las motivaciones, incentivos o intereses de los agentes o actores sociales para explicar los problemas según sus contextos (Bartelds, Savenije y Van Boxtel, 2020; Henríquez *et al.*, 2018; Siegfried *et al.*, 2010; Smith, 2014). Finalmente, al determinar las causas de los problemas, los estudiantes juzgan procesos o eventos desde su significancia histórica (Barton, 2005; Seixas, 2017; VanSledright y Limón, 2006) y utilizan metaconceptos y abstracciones en sus explicaciones (Siegfried *et al.*, 2010; Van Drie y Van Boxtel, 2008).

La *interpretación de las fuentes sobre problemas sociales* implica que las personas se aproximan críticamente a las evidencias de información a través de la identificación de perspectivas y la diferenciación de hechos y puntos de vista (Reisman, Brimesk y Hollywood, 2019). Esto también implica incorporar evidencias para fundamentar una posición, corroborando y reformulando las fuentes y reconociendo sus autorías (Breakstone, Smith y Wineburg, 2013; Henríquez *et al.*, 2018; McGrew, Breakstone, Ortega, Smith *et al.*, 2018; Seixas, 2017).

Finalmente, la dimensión *argumentación de propuestas* permite fundamentar una posición respecto de un problema social (National Council for the Social Studies, 2013; Ross, 2019; Seixas, 2017) y elaborar propuestas o alternativas de decisión pertinentes (Gerwin, 2003). Estas ideas deberían reconocer contextos locales, regionales y globales (National Council for the Social Studies, 2013), riesgos, el principio de escasez (Siegfried *et al.*, 2010), críticas a la desigualdad (Pedrozo, 2011) y promoción del bienestar común de la sociedad (Niepel, Mustafić, Greiff y Roberts, 2015); también deberían estar basadas en un enfoque de derechos, democrático, en perspectiva de equidad y una ética centrada en el bien común (Levy, 2018; Ross, 2019; Seixas, 2017). Asimismo, las propuestas requieren ser

concretas, lo que implica reconocer en ellas el rol que los actores sociales desempeñan en su implementación (Evans y Saxe, 1996).

La propuesta planteada a continuación presenta un instrumento que pretende evaluar el pensamiento social en el aula, junto a las evidencias psicométricas sobre la validez de sus puntuaciones.

Método

Desde un enfoque cuantitativo y en la perspectiva de un análisis psicométrico, se describe la construcción y el proceso de recolección de evidencias de validez (American Educational Research Association, American Psychological Association y National Council on Measurement in Education, 2014) de las puntuaciones de un instrumento diseñado para evaluar el pensamiento social de los estudiantes egresados de la educación secundaria. En una primera etapa se describe el diseño del instrumento y, en la segunda, se presenta el proceso de recolección de evidencias de validez.

El estudio espera reportar evidencias favorables respecto del contenido, fiabilidad entre jueces y consistencia interna de las puntuaciones, además de evidencias sobre la estructura interna de cuatro dimensiones que componen el constructo, con el fin de retroalimentar el pensamiento social. Asimismo, se espera observar relaciones positivas entre las puntuaciones del instrumento y el rendimiento académico de las y los estudiantes en el contexto escolar; aunque no se espera observar relaciones con el género.

Etapas 1. Diseño del instrumento

El instrumento se diseñó siguiendo el modelo del Bear Assessment System (BAS) (Wilson, 2005), a través de las siguientes etapas: *a)* definir los cuatro mapas de constructo para cada una de las dimensiones de pensamiento social (contextualización, explicación, interpretación y argumentación de propuestas); *b)* diseñar la tarea de desempeño; *c)* diseñar la rúbrica; y *d)* implementar el modelo de medición para recopilar evidencias de validez.

Esta primera versión del instrumento fue desarrollada por un equipo de especialistas en didáctica de las ciencias sociales, a partir de una propuesta teórica (Gutiérrez y Pagés, 2018; Pagés, 2002) y de evidencia empírica sobre las habilidades de las dimensiones y de las habilidades que lo componen. El contenido de las cuatro primeras etapas fue analizado a través del juicio de expertos, en este caso, tres especialistas externos al equipo de investigación. El modelo de medición fue implementado por el equipo de autoras.

Etapa 2. Recopilación de evidencias de validez empíricas

Muestra

El instrumento fue aplicado a una muestra por conveniencia de 175 estudiantes egresados de educación media en Chile, durante la primera semana de clases en una universidad pública y selectiva. Todos estudiaban carreras de pedagogía, donde 74% fueron mujeres y 36% varones, con una edad entre 17 y 20 años, y de diversos contextos socioeconómicos. En ese sentido, los resultados no esperan ser representativos de la población de egresados de educación secundaria chilena.

Aplicación y corrección

Tras un proceso de capacitación, los aplicadores siguieron un protocolo estandarizado para la implementación del instrumento utilizando la plataforma Moodle (versión 3.7). Al inicio de la aplicación, se leyó y firmó el consentimiento informado por parte de los participantes. Posteriormente, los estudiantes contaron con 90 minutos para responder el instrumento descrito más adelante.

La revisión se realizó con cinco correctores capacitados, quienes corrigieron en forma simultánea después de un periodo de calibración. Un 30% de los discursos fueron corregidos por dos jueces independientes.

Análisis de evidencias de validez

Para analizar las puntuaciones se exploró su fiabilidad, evidencias de validez en términos de su estructura interna y su relación con otras variables. La mayoría de los análisis fueron realizados a través del programa Rstudio versión 3.6.0, aunque el análisis de ítems se realizó con el software Berkeley Assessment System Software (BASS) (Fisher y Wilson, 2019).

Fiabilidad del instrumento

Después de constatar una distribución aproximadamente normal del puntaje total del instrumento, se realizó un análisis de fiabilidad, a partir de un examen de consistencia interna a través de alfa de Cronbach (Raykov, 1998) para cada una de las dimensiones del instrumento. Adicionalmente, se hizo la estimación de la fiabilidad a través del indicador *expected a-posteriori* (EAP), debido a su mayor pertinencia en el contexto de los análisis de teoría de respuesta al ítem (IRT, por sus siglas en inglés) (Adams, 2005).

Finalmente, el análisis de kappa de Cohen (Embretson, 2010) se utilizó para analizar la fiabilidad entre jueces en el 30% (N = 52) de los casos con doble corrección.

Estructura interna del instrumento

Análisis de ítems

Después de analizar la estructura de la matriz de correlaciones se analizó el comportamiento empírico de los ítems que componen las cuatro dimensiones del pensamiento social. Se llevó a cabo un análisis IRT con base en un modelo de crédito parcial por la naturaleza politómica de las categorías de respuesta del instrumento (Wilson, 2005). Una vez realizada la comparación de los modelos multidimensional y unidimensional, se analizó la dificultad de los ítems y el indicador rho de Spearman de cada dimensión. Adicionalmente, se elaboró un mapa de Wright con el fin de reconocer la relación entre la dificultad de los ítems y la probabilidad de los sujetos de responder en los distintos niveles de logro (Wilson, 2005).

Análisis factorial

Con el fin de analizar la estructura interna del instrumento con mayor profundidad, se aplicó un análisis factorial confirmatorio (AFC). El ajuste del modelo fue estudiado por medio de los índices RMSEA, CFI, TLI, SRMS (Hooper, Coughlan y Mullen, 2008). Para el AFC se exploró un modelo de uno y cuatro factores; mientras que el unidimensional representa la hipótesis nula, y el de cuatro dimensiones corresponde a la hipótesis propuesta por el modelo teórico.

Relación con género y rendimiento académico de los estudiantes

Después de examinar la estructura interna del instrumento, se implementó un modelo estructural para estudiar la relación entre las dimensiones del pensamiento social y algunas variables relevantes, como el *ranking* de notas (calificaciones) de educación secundaria y el género de los estudiantes. El *ranking* de notas utilizado en el proceso nacional de admisión fue considerado como un proxy del rendimiento académico de los estudiantes, ya que, a diferencia de las calificaciones en sí y de la Prueba de Selección Universitaria, dicho *ranking* de educación secundaria considera el desempeño de los estudiantes en relación con su contextos educativos y oportunidades de aprendizaje (Larroucau, Ríos y Mizala, 2015). El género, por su parte,

fue incorporado debido a la evidencia respecto de la brecha favorable a niñas y mujeres jóvenes en las habilidades de escritura académica y lengua (Cervini, Dari y Quiroz, 2015; Contreras-Salinas, Olavarría Aranguren, Celedón Bulnes y Molina Gutiérrez, 2020). Una relación positiva entre ser mujer y tener un mayor desarrollo del pensamiento social no constituiría evidencia a favor de la validez del instrumento.

Se analizó la relación de las puntuaciones del instrumento con género y el *ranking* de notas a través de un modelo de ecuaciones estructurales (Schreiber, Nora, Stage, Barlow *et al.*, 2006). Para estudiar el ajuste del modelo, se utilizaron los índices RMSEA, CFI, TLI, SRMS (Hooper, Coughlan y Mullen, 2008).

Resultados

Las evidencias presentadas a continuación aportan antecedentes a favor y de mejora respecto de la validez de las puntuaciones del instrumento diseñado para evaluar el pensamiento social de los estudiantes desde una perspectiva interdisciplinaria del aprendizaje de las ciencias sociales. Los resultados se presentan en dos secciones: en primer lugar, se describe el diseño del instrumento respecto de los tres primeros elementos del modelo BAS y luego se presentan los resultados del cuarto elemento, el análisis psicométrico.

Eta 1. Diseño del instrumento

El instrumento consistió en un problema socialmente relevante (Pagés y Santisteban, 2011) presentado en una carpeta de fuentes de información de diversa naturaleza (noticias, extractos de libros de historia, gráficos, etc.) y una consigna de instrucciones que solicitaban la elaboración de un discurso escrito usando el computador. Para evaluar el discurso hecho por los estudiantes, se confeccionó una rúbrica analítica con 18 criterios relacionados con el constructo y graduados entre 0 a 3 (anexo 1).

El instrumento consta de tres partes: en la primera, los estudiantes leen el contexto y el problema socialmente relevante presentado (figura 1). Posteriormente, deben leer diferentes fuentes de relacionadas con el problema y extraer información considerada relevante. Finalmente, el instrumento pide a los estudiantes escribir su discurso a través de una breve consigna que orienta sobre los elementos que debe considerar el discurso y el tiempo para desarrollarlo (figura 2).

FIGURA 1

Planteamiento del problema socialmente relevante

Lee atentamente la siguiente situación

Durante los últimos 5 años, Chile ha recibido una cantidad de población migrante que supera con creces la de cualquier otro periodo de su historia republicana. Diversos conflictos políticos, sociales y económicos en países latinoamericanos han movilizado a miles de personas al país, ante la expectativa de encontrar mejores condiciones de vida.

Debido a este aumento explosivo, han surgido voces a favor de desarrollar una Nueva Ley de Migraciones que mejore los procesos de ingreso al territorio, con el argumento de que la institucionalidad vigente no es suficiente. Sin embargo, otro grupo de personas está en contra de la reforma, porque argumentan, podría favorecer abusos y vulnerar los derechos de las y los migrantes.

En este sentido, ¿por qué si se plantea que la Nueva Ley de Migraciones mejorará el proceso migratorio, existen personas en contra de su implementación?

La controversia planteada está generando gran debate en la comunidad universitaria y ha sido elegida el tema central para inaugurar el año académico de la carrera. Debido a esto deberás elaborar un discurso que aborde la controversia.

Para desarrollar el discurso debes seguir las instrucciones planteadas más adelante.

Fuente: instrumento diseñado para esta investigación.

FIGURA 2

Instrucciones para desarrollo de la tarea

Escribe tu discurso utilizando una estructura de inicio, desarrollo y cierre.

Recuerda que la controversia es: ¿por qué si se plantea que la Nueva Ley de Migraciones mejorará el proceso migratorio, existen personas en contra de su implementación?

Considera los siguientes elementos en tu discurso:

- a) La presentación del tema.
- b) Exposición de argumentos explicativos.
- c) Fundamentar los argumentos utilizando fuentes de información.
- d) Propuestas para abordar la situación.

Sugerencias:

1. Antes de responder, haz un esquema o toma apuntes que te permitan organizar tu discurso.
Para eso utiliza una hoja en blanco en tu casa.
2. Para realizar el discurso cuentas con 60 minutos (si demoraste 30 en la lectura).
3. Debes utilizar entre 20 y 40 líneas.

Fuente: instrumento diseñado para esta investigación.

Las respuestas de los estudiantes demostraron cumplir con la idea de elaboración de un discurso, con 30 líneas en promedio y basados en la estructura indicada en las instrucciones. En la figura 3 se presenta un extracto de la respuesta de un estudiante.

FIGURA 3

Extracto de respuesta

Estimados lectores,

Como todos sabemos Chile es el principal destino de migrantes a nivel latinoamericano, atrayente por sus características físicas, prosperidades económicas y de crecimiento. Tanto así que dentro de las estadísticas de los últimos años un 92,4% de las personas que viven en Chile son inmigrantes (...)

De esta manera se discrimina a un extranjero por pensar que solo su propósito de venir al país es a delinquir, no obstante, sabemos que hace más de un siglo los inmigrantes viene por oportunidades de trabajo o a establecer una nueva vida ya que no tenían las condiciones apropiadas en su original país

(...)

Fuente: Respuesta de un estudiante participante en esta investigación.

Cada uno de los discursos fue evaluado con la rúbrica analítica (anexo 1). En la dimensión *contextualización* fue evaluada la localización tanto temporal como espacial, actores y ámbitos. En la *explicación* se evaluó multicausalidad, cadena causal, empatía, conceptualización, significancia histórica y duración. Para la *interpretación de fuentes* los criterios fueron perspectiva, corroboración, forma de la fuente y referenciación. Y para el caso de la *argumentación de propuestas*, los criterios fueron acción de los agentes, propuesta, enfoque de derechos y posicionamiento.

Tanto el diseño de la tarea como los criterios de la rúbrica fueron analizados en su contenido por tres especialistas en didáctica de las ciencias sociales, reunidos en sesiones de discusión grupal. Ellos coincidieron respecto de la capacidad del instrumento de evaluar el pensamiento social y discriminar usando los criterios propuestos.

Etapa 2. Recopilación de evidencias de validez empíricas

Consistencia del instrumento

Los estadísticos descriptivos del instrumento muestran que los puntajes totales de los estudiantes van desde 5 hasta 87% de logro. La consistencia interna del instrumento es aceptable, aunque los resultados son mejores en su estimación a través del método EAP para todas las dimensiones (tabla 1) que utilizando alfa de Cronbach (tabla 2); este último análisis permite identificar que la dimensión *contextualización* presenta una fiabilidad menor a la óptima que requiere ser revisada a la luz de las demás evidencias.

TABLA 1

Fiabilidad EAP

Contextualización	Explicación	Interpretación de fuentes	Argumentación de propuesta
0.86	0.89	0.80	0.82

Fuente: elaboración propia.

TABLA 2

Fiabilidad por alfa de Cronbach

Contextualización	Explicación	Interpretación de fuentes	Argumentación de propuesta
0.61	0.86	0.80	0.78

N=175.

Fuente: elaboración propia.

La confiabilidad entre jueces presenta resultados favorables. Para los 51 casos evaluados por doble juez, el índice de acuerdo total alcanza el 92%. Adicionalmente, se observa un kappa de Cohen donde todos los ítems superan 0.81, aunque uno de ellos (significancia histórica) presenta una estimación de 0.73, en un intervalo de confianza que va desde 0.56 a 0.90 (tabla 3). Esto alerta sobre la necesidad de revisar el ítem a la luz de otras evidencias.

TABLA 3

*Intervalo de confianza kappa de Cohen del ítem significancia histórica**

	Inferior	Estimación	Superior
Kappa no ponderado	0.56	0.73	0.90
Kappa ponderado	0.78	0.78	0.78

* Con 51 casos evaluados a doble juez.

Fuente: elaboración propia.

Evidencias de validez sobre la estructura interna del instrumento

Al analizar la matriz de correlaciones entre dimensiones (tabla 4), se identificaron relaciones moderadas entre las del instrumento, excepto por la alta correlación (0.8) entre las dimensiones *contextualización* y *explicación*. Sobre esos antecedentes, se continuó el análisis de la estructura interna del instrumento de evaluación, a través de dos tipos de análisis: de ítems y AFC.

TABLA 4

*Correlaciones entre dimensiones**

	Contextualización	Explicación	Interpretación de fuentes	Argumentación de propuestas
Contextualización	1			
Explicación	0.80	1		
Interpretación de fuentes	0.46	0.40	1	
Propuestas	0.40	0.52	0.34	1

* $p = 0$ para todas las correlaciones, indica que las correlaciones son estadísticamente significativas.

Fuente: elaboración propia.

Análisis de ítems

Con el fin de analizar la estructura interna del instrumento, se comparó el modelo teórico de cuatro dimensiones con el unidimensional, usando el modelo IRT (Wilson, 2005).

El análisis rechaza la hipótesis planteada por el modelo unidimensional, lo que permitió considerar el teórico de cuatro dimensiones, observándose diferencias estadísticamente significativas entre ambos modelos ($p = 0$, $\chi^2 = 253$, $gl = 9$), con una ventaja en todos los índices de ajuste por parte del modelo multidimensional (tabla 5). En ese sentido, fue posible sostener que el pensamiento social es un constructo compuesto por cuatro dimensiones.

TABLA 5

*Comparación del modelo unidimensional y multidimensional**

Modelo	loglike	Deviance	Npars	Nobs	AIC	BIC	SRMR	pmaxX2
Multidimensional (4 dimensiones)	-2982	5 963	82	175	6 127	6 387	0.089	1.2e-12
Unidimensional	-3108	6 216	73	175	6 362	6 593	0.124	5.8e-07

* $p = 0$, $\chi^2 = 253$, $gl = 9$.

Fuente: elaboración propia.

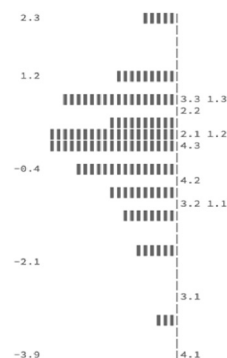
Al profundizar en el análisis del modelo IRT multidimensional desde los mapas de Wright de cada dimensión (figura 4) y la estimación de rho de Spearman fue posible identificar que la distribución de las dificultades de los ítems se comporta de acuerdo con la hipótesis planteada. Los estudiantes y los ítems se distribuyen en todo el espacio de respuesta de cada dimensión. En las cuatro dimensiones, el índice de relación entre la hipótesis y los resultados es alta (rho de Spearman > 0.71). En la dimensión de *interpretación de fuentes* el nivel de correlación es aún mayor (rho de Spearman > 0.91). En ese sentido, la mayoría de los ítems en cada dimensión se distribuye en forma coherente con la hipótesis planteada, y las categorías de logro esperadas en los ítems coinciden con la dificultad empírica presentada. Sin embargo, la dimensión *contextualización* presentó dificultades respecto de su hipótesis inicial. El ítem localización geográfica (GEOL) omitió la representación de uno de los niveles de logro, debido a que ninguno de los estudiantes demostró su desempeño en el nivel 0. Adicionalmente, es posible reconocer que casi todos los ítems demuestran niveles de ajuste apropiados ($0.75 > \text{infit} > 1.33$), excepto los correspondientes a significancia histórica ($\text{infit} > 1.33$) y conceptualización ($\text{infit} > 1.33$), ambos estadísticamente significativos

($p \leq 0.05$). El primero de ellos presentó dificultades en el nivel 1 y el segundo en el nivel 2 de la rúbrica de evaluación.

FIGURA 4

Mapa de Wright por dimensión

A. Dimensión contextualización

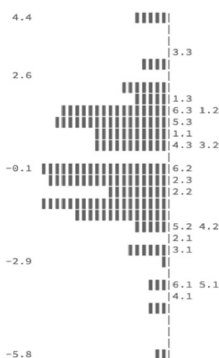


Each  represented up to 1.45 cases.
Generated using WLE estimates.

Ítems

1. Localización temporal
2. Localización espacial
3. Ámbitos
4. Agentes

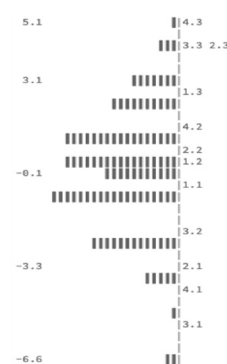
B. Dimensión explicación



Each  represented up to 1 case.
Generated using WLE estimates.

- | | |
|----------------------------|--------------------|
| 1. Significancia histórica | 5. Multicausalidad |
| 2. Conceptualización | 6. Causalidad |
| 3. Continuidad y cambio | |
| 4. Empatía | |

C. Dimensión interpretación de fuentes

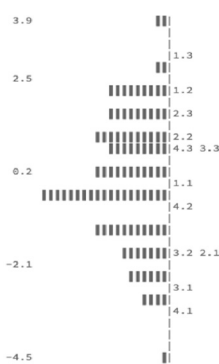


Each  represented up to 1.65 cases.
Generated using WLE estimates.

Ítems

1. Referenciación
2. Corroboración
3. Forma de la fuente
4. Perspectiva

D. Dimensión argumentación de propuestas



Each  represented up to 1.75 cases.
Generated using WLE estimates.

1. Acción de los agentes
2. Propuesta
3. Enfoque de derechos
4. Posicionamiento

Fuente: elaboración propia.

Sobre la dificultad de los ítems, el más fácil de lograr fue el de empatía (-1.78), mientras que el más difícil fue acción de los agentes (1.81). Esto implica que para alcanzar los distintos niveles del ítem empatía no es necesario tener un nivel de pensamiento social muy alto, mientras que para lograr los niveles del ítem acción de los agentes se requiere un alto desarrollo de la habilidad.

Análisis factorial confirmatorio

En el caso de la estructura factorial también se probó un modelo unidimensional y otro de cuatro factores (Gutiérrez y Pagés, 2018). En ambos casos el análisis demuestra que el índice chi cuadrado no es significativo ($p = 0.000$) y que por tanto se rechaza la hipótesis nula de una incorrecta especificación de los modelos. Adicionalmente, al analizar el ajuste de ambos modelos fue posible identificar que los índices de bondad de ajuste RMSEA, CFI, TFI y SMRM se encuentran fuera de los rangos esperados (tabla 6). Estos antecedentes, junto al examen de las cargas factoriales del modelo teórico, la fiabilidad de las dimensiones, la correlación entre dimensiones y la representación de la matriz de correlaciones entre ítems podrían sugerir que el modelo teórico inicial no está logrando diferenciar entre las dimensiones *contextualización* y *explicación*.²

TABLA 6

*Índices de ajuste. Modelo factorial confirmatorio**

			Modelo unidimensional	Modelo 4 dimensiones	Modelo 3 dimensiones	Modelo 4 dimensiones modificado
Modelo con relación entre factores	RMSEA	< 0.08 -0.06	0.139	0.108	0.109	0.080
	CFI	> 0.95	0.673	0.812	0.804	0.900
	TLI	> 0.95	0.629	0.777	0.772	0.877
	SRMR	< 0.08	0.104	0.080	0.082	0.072

*N=175.

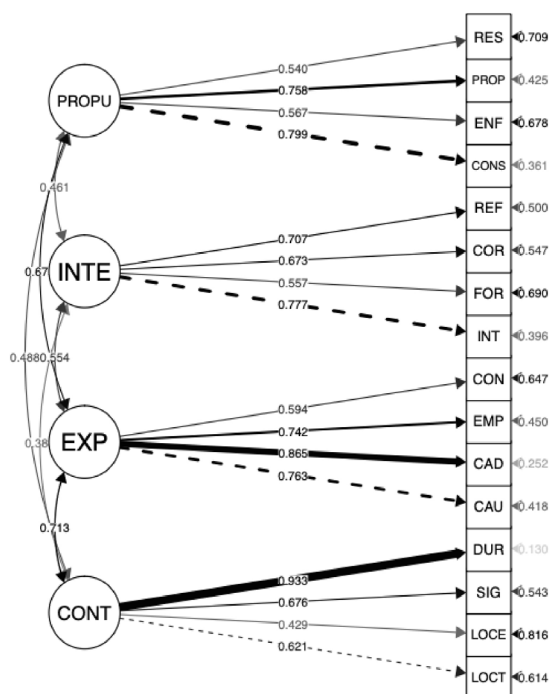
Fuente: elaboración propia.

Profundizando en los resultados, se observó que algunos de los criterios de la dimensión *explicación* estaban describiendo la temporalidad del

problema, más que dando cuenta de las causas que lo generan. Por esto se exploró un nuevo modelo de cuatro dimensiones “modificado”. La nueva hipótesis implicó reconstruir la dimensión *contextualización* manteniendo los criterios de localización temporal y localización geográfica, pero incorporando los criterios de significancia histórica y de duración (originalmente de la dimensión *explicación*). Se testeó este modelo debido a que la revisión de la rúbrica permitió identificar que ambos criterios no estaban permitiendo indagar las explicaciones de los estudiantes, sino más bien elementos contextuales que describían los problemas en sus relatos. A partir de dicha reagrupación, el modelo de cuatro dimensiones modificado (figura 5) permitió identificar una mejora considerable en los índices de ajuste (ver tabla 6), aunque todavía sin lograr niveles de ajuste óptimos.

FIGURA 5

Representación gráfica modelo de cuatro dimensiones ajustado



Fuente: elaboración propia.

Evidencias de validez sobre relación con el género y el rendimiento académico

Contando con cierta evidencia respecto de la estructura multidimensional del modelo ajustado, se realizó un análisis basado en el modelo de ecuaciones estructurales, para relacionar las dimensiones del pensamiento social con el género de los estudiantes y el *ranking* de notas de la educación secundaria de los estudiantes. Se hipotetizaba una relación estadísticamente significativa y positiva de los puntajes de las dimensiones del pensamiento social con el *ranking* de notas y una relación no estadísticamente significativa con el género. En ese sentido, se esperaba que alumnos de rendimiento escolar relativamente superior tuvieran mejor desempeño en el instrumento diseñado. Al mismo tiempo, se esperaba que las mujeres tuvieran mejores puntuaciones que los varones.

Los índices de ajuste del modelo se consideran aceptables ($\chi^2 = 212,48$, $p = 0.000$, $gl = 114$) (tabla 7), aunque dos de ellos están levemente por debajo de los parámetros esperados.

TABLA 7

*Índices de ajuste Modelo de Ecuaciones Estructurales**

			Modelo probado
Índices de ajuste	RMSEA	< 0.08 -0.06	0.071
	CFI	> 0.95	0.897
	TLI	> 0.95	0.878
	SRMR	< 0.08	0.071

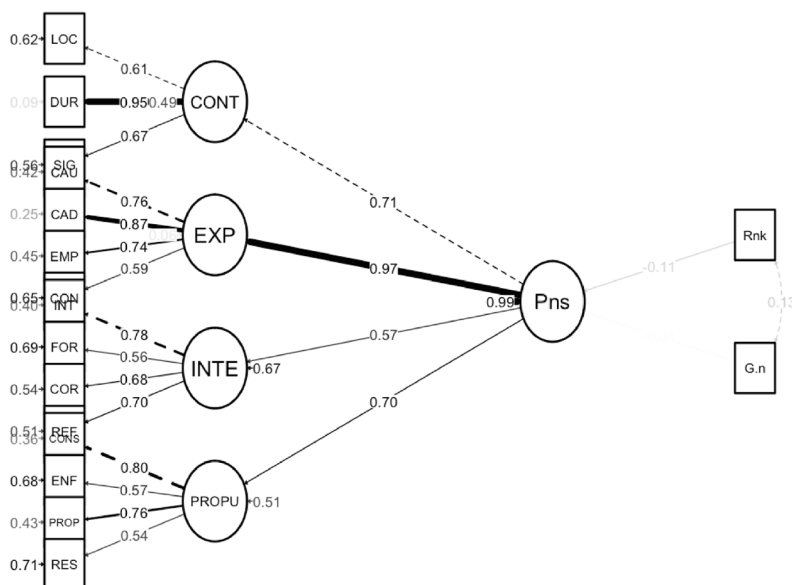
* N = 175. Método de estimación oblicua.

Fuente: elaboración propia.

Respecto del modelo estructural (figura 6), es posible inferir que la relación entre el pensamiento social y el puntaje *ranking* de notas (Rkg) de los estudiantes no es estadísticamente significativa ($p > 0.05$). Tampoco existe evidencia de una relación estadísticamente significativa entre los puntajes del constructo y el género (G.n) de los estudiantes ($p > 0.05$).

FIGURA 6

Representación gráfica del modelo empírico



Fuente: elaboración propia.

Discusión y conclusiones

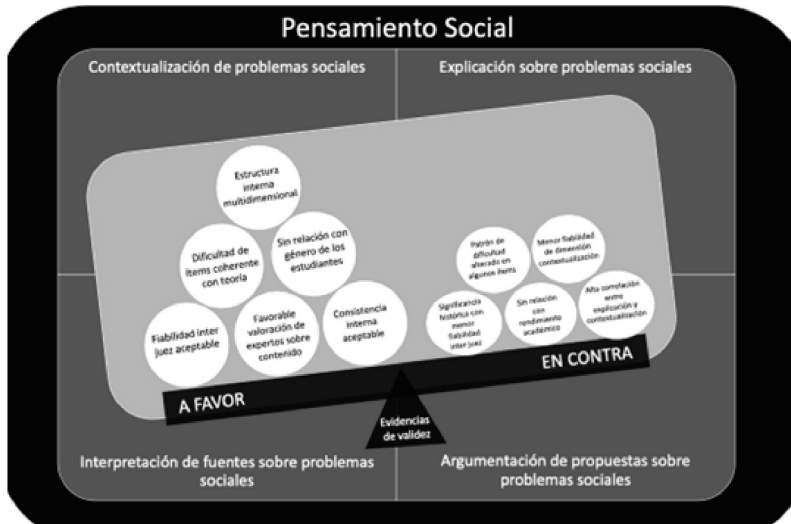
La enseñanza de las ciencias sociales se enfrenta hoy a importantes obstáculos, debido en gran medida a que las prácticas de enseñanza y evaluación del profesorado continúan centradas en el desarrollo de aprendizajes reproductores de información (Torrez y Claunch-Lebsack, 2014). Ante estos obstáculos, la investigación ha avanzado en dos perspectivas: la disciplinaria y la interdisciplinaria del aprendizaje desde los estudios sociales.

La perspectiva interdisciplinaria de los estudios sociales se ha propuesto avanzar en promover el pensamiento social en los estudiantes desde la integración de habilidades de las ciencias sociales (Gutiérrez y Pagés, 2018). Sin embargo, las investigaciones sobre el qué y cómo medir y evaluar el aprendizaje de las ciencias sociales siguen centrándose en una perspectiva disciplinaria (pensamiento histórico, por ejemplo) de forma específica y continúan presentando desafíos sobre la validez y fiabilidad de las puntuaciones (Smith, Breakstone y Wineburg, 2018; Wainer, 2011; Grant, Gradwell y Cimbricz, 2004).

Este estudio informa sobre el diseño de un instrumento de evaluación capaz de medir y retroalimentar el aprendizaje en el aula según los especialistas siguiendo las tres primeras etapas del método BAS (Wilson, 2005). Además, presenta evidencias sobre la validez de sus puntuaciones, aportando antecedentes para el desarrollo de evaluaciones de desempeño capaces de orientar a los estudiantes de manera integrada respecto de sus habilidades para contextualizar, explicar, interpretar y argumentar sobre problemas sociales. Esto se hace a partir de las puntuaciones asignadas a las respuestas del instrumento por medio de un modelo IRT, tal como se describe en la cuarta y última etapa del método de BAS. Los hallazgos a favor del argumento sobre validez de las puntuaciones para la evaluación en aula (Kane, 2020) permiten avanzar hacia la consolidación de un referente evaluativo más coherente con los propósitos de la enseñanza de las ciencias sociales (Percy, 2017; Torrez y Claunch-Lebsack, 2013). La figura 7 sintetiza esos principales logros, pero también presenta las evidencias débiles, con el fin de orientar su mejora.

FIGURA 7

Evidencias a favor y en contra de la validez de las puntuaciones del instrumento



Fuente: elaboración propia.

Las evidencias sustentan un constructo multidimensional, con grados aceptables de consistencia y fiabilidad entre jueces, que logra identificar niveles de logro coherentes con las hipótesis definidas, y relaciones plausibles con otras variables. Sin embargo, a pesar de que las evidencias son positivas a la luz de hallazgos previos (Wainer, 2011; Grant, Gradwell y Cimbricz, 2004), existen desafíos que deben ser atendidos.

La propuesta diseñada para evaluar el pensamiento social

El instrumento consiste en una situación de desempeño escrita basada en un problema socialmente relevante. El ensayo elaborado por el estudiante es analizado a través de una rúbrica que permite identificar las habilidades de pensamiento social mediante 18 criterios derivados de cuatro dimensiones: la *contextualización*, la *explicación*, la *interpretación de las fuentes* y la *argumentación de propuestas* (Gutiérrez y Pagés, 2018; Pipkin y Sofía, 2005). El análisis de expertos evaluó positivamente las posibilidades del instrumento y la rúbrica de medir el pensamiento social con el fin de retroalimentar su aprendizaje en el aula de ciencias sociales.

Confiabilidad aceptable con posibilidades de mejora

A diferencia de los resultados de fiabilidad presentados por Wainer (2011), se observa que el instrumento presenta puntuaciones estables, tanto en su consistencia interna como en el acuerdo de los jueces. Esto indicaría que los resultados son consistentes y que no dependen de quién utilice el instrumento. Además, el modelo multidimensional presenta una fiabilidad aceptable para cada una de las dimensiones. Se observa, sin embargo, menor nivel de fiabilidad para la correspondiente a *contextualización*, lo que podría estar relacionado con que esta dimensión requiere movilizar otras habilidades como la escritura o la lectura, más que en las otras (De La Paz *et al.*, 2012).

Por otra parte, se debe precisar la descripción de los niveles de logro del ítem significancia histórica, ya que presenta niveles de acuerdo entre jueces menores al resto de los ítems. Aunque los resultados respecto de confiabilidad son más altos que aquellos de otras mediciones (Wainer, 2011), es posible reconocer posibilidades de mejora. Por ejemplo, el nivel dos de la rúbrica en este ítem indica que se incorporan eventos o procesos históricos que “no tienen un rol clave en la explicación”. Esta falta de precisión en la descripción podría estar dando espacio a interpretaciones particulares

de los jueces, lo que podría explicar la falta de acuerdo y diferencias entre puntuaciones.

Un constructo multidimensional y el desafío de la dimensión *contextualización*

Los análisis realizados reconocen la estructura multidimensional del pensamiento social, pero también sugieren posibilidades de mejora. La dimensión *contextualización* se relaciona de forma estrecha con la de *explicación*, lo que podría implicar que los ítems que las componen no están diferenciando la intencionalidad de ambos factores por separado (contextualizar y explicar, respectivamente).

La dimensión *contextualización* pretende reconocer cómo los estudiantes describen los antecedentes del problema social, considerando su marco temporal y espacial, junto con los ámbitos y actores en que se manifiesta el problema (Gutiérrez y Pagés, 2018); mientras que la correspondiente a *explicación* busca reconocer las causas y consecuencias que los estudiantes despliegan para explicar los problemas sociales, reconociendo las motivaciones de los actores, la nominalización de los procesos y el reconocimiento de ciertos eventos históricos. A pesar de que algunos autores (Lee, 2009) plantean que las habilidades de explicación podrían contener las de contextualización, los análisis demuestran que el modelo que unifica ambas dimensiones no logra mejores índices de ajuste que el modelo de cuatro. Es por esto que próximas investigaciones deberían profundizar en la examinación de estas dimensiones.

Sin embargo, el análisis realizado podría aportar antecedentes relevantes para diferenciar ambas dimensiones. Por ejemplo, los ítems agentes (de la dimensión *contextualización*) y empatía (de *explicación*) requieren ser precisados. Cuando la rúbrica aborda el ítem agentes indica en el nivel máximo que “incorpora actores, agrupaciones y/o instituciones relacionadas de forma directa e indirecta en la descripción de la problemática, que intervienen y toman decisiones sobre la problemática”. No obstante, esta descripción se relaciona directamente con la habilidad de empatía, ya que implica reconocer las motivaciones y posiciones de los actores para desarrollar una determinada acción social. Esto último es también parte de las explicaciones que realizan los estudiantes sobre los problemas sociales. En ese sentido, una próxima versión del instrumento debería diferenciar con mayor claridad el ítem agentes del de empatía.

Ítems por mejorar:

localización geográfica y conceptualización

Se identificaron otras problemáticas en algunos ítems de las dimensiones *contextualización y explicación*. Como se comentó previamente, el puntaje 0 del ítem localización geográfica no fue identificado en ninguno de los discursos de los estudiantes. Esto implica la necesidad de reconfigurar dicha descripción, debido a que posiblemente está representando niveles poco plausibles para la edad de los estudiantes.

En el caso del ítem conceptualización se reconocen problemas en el patrón de dificultad de los ítems. Entre las posibles causas se puede reconocer que la rúbrica no diferencia entre los niveles dos y tres la utilización de los denominados metaconceptos o conceptos de segundo orden que proponen Van Drie y Van Boxtel (2008), lo que podría estar interfiriendo en la distribución de la dificultad de ambos niveles de logro.

Coherencia entre la construcción teórica y la evidencia empírica

A pesar de que se observaron algunas necesidades de mejora, es posible reconocer logros relevantes al definir los criterios para evaluar el pensamiento social. Además de la consistencia interna y la confiabilidad entre jueces que se reporta, es relevante mencionar que la evidencia apoya la conceptualización del pensamiento social como un constructo multidimensional.

Adicionalmente, se identifica que la mayoría de los ítems presentan una distribución de la dificultad coherente con la hipótesis diseñada. Esto implica que los niveles de logro que se esperaban fueran de menor dificultad efectivamente así se presentaron. Lo mismo sucede con los niveles de logro de mayor dificultad. Es decir, se observan evidencias a favor de las puntuaciones del instrumento.

La inexistente relación entre pensamiento social, género y rendimiento académico

En el análisis fue posible identificar que el género no tiene una relación estadísticamente significativa con las puntuaciones del test. Si bien este es un antecedente relevante, es importante corroborar estas conclusiones en otros estudios, debido a la amplia literatura disponible con respecto a las brechas de género que favorecen a las mujeres en el área de la lengua

y las humanidades (Cervini, Dari y Quiroz, 2015; Contreras-Salinas *et al.*, 2020).

Adicionalmente, la inexistencia de relaciones estadísticamente significativas entre las puntuaciones del instrumento y las calificaciones de los estudiantes en la escuela podrían explicarse por las dinámicas reproductivas y memorísticas del aprendizaje que están ampliamente documentadas en el sistema escolar de diversos países (Torrez y Claunch-Lebsack, 2013). Es sabido que esas prácticas evaluativas “premian” en sus calificaciones a estudiantes capaces de responder a esa dinámica, más que a aquellos que logran demostrar un pensamiento más complejo (Pearcy, 2017). En ese sentido, los hallazgos subrayan la importancia de continuar explorando la potencialidad del pensamiento social para promover aprendizajes significativos.

Desafíos y proyecciones

Los principales desafíos para el desarrollo del instrumento se encuentran en lograr una estimación más precisa de las dimensiones *contextualización* y *explicación*. Sin embargo, también es necesario alcanzar conclusiones más definitivas respecto de la relación de las puntuaciones de pensamiento social con otras variables relevantes, tales como el nivel sociocultural de los estudiantes. Además, se deben analizar muestras que permitan realizar inferencias generalizables a la población escolar.

En el futuro se espera avanzar hacia modelos de medición con mejores índices de ajuste y que permitan explorar vínculos con otras variables exógenas. En especial, se debería analizar la relación entre el pensamiento social y la formación ciudadana de los estudiantes, además de poder diferenciar estos aprendizajes de habilidades mediadoras como la lecto-escritura académica. Adicionalmente, es necesario avanzar en investigaciones que permitan explorar nuevas evidencias de validez del instrumento, tales como sus consecuencias y usos en contextos de aula (American Educational Research Association, American Psychological Association y National Council on Measurement in Education, 2014).

Este estudio abre nuevos caminos y entrega posibles directrices que orientan al desarrollo de procesos evaluativos auténticos que permitan retroalimentar el aprendizaje de las ciencias sociales en el aula. Esto podría permitir impulsar el aprendizaje en un subsector que, aunque privilegiado para formar ciudadanía, ha subutilizado su potencialidad.

ANEXO

Rúbrica de evaluación del pensamiento social

Criterios	Nivel 3	Nivel 2	Nivel 1	Nivel 0
1. Localización temporal	Utiliza marcas temporales disciplinares para situar el problema (siglos, décadas, años, épocas, periodos, actualidad), tanto en el presente y respecto de procesos del pasado UB3 <i>Durante la última década...</i>	Sitúa el problema en el presente y con referencias al pasado, pero utiliza marcas temporales de naturaleza coloquial (antes, posterior, ahora, hoy) UBI2 <i>Hace tiempo que...</i>	Utiliza marcas temporales para situar el problema (siglos, décadas, años, épocas, periodos, en la actualidad), pero solo en una mirada del presente UBI1 <i>En la actualidad...</i>	Sitúa el problema sólo desde el presente, utilizando marcas temporales de naturaleza coloquial UBI0 <i>Hoy...</i>
2. Localización espacial	Sitúa la problemática en el espacio, utilizando categorías de posición lexicales (Chile, Santiago, en Europa) e indexicales (al sur de, el país), relacionando distintas escalas territoriales con categorías de posición relativas GEO3 <i>Comunas de la zona sur de la región, como Puente Alto.</i>	Sitúa la problemática en el espacio, utilizando categorías lexicales (Chile, Santiago, en Europa) o indexicales (al sur de, el país, la región, las poblaciones), con existencia de marcas de escalas geográficas con categorías de posición relativas GEO2 <i>En Chile, la relación con América...</i>	Sitúa la problemática en el espacio, utilizando categorías indexicales (al sur de, el país, la región, las poblaciones) y/o lexicales (Chile, Santiago, en Europa), pero que no precisan escalas territoriales con categorías de posición relativas GEO1 <i>En el lugar...</i>	Solo utiliza categorías de posición indexicales sin escala territorial (ahí, allá, acá) para situar el territorio estudiado. GEO0 <i>Aquí...</i>
3. Problematicación	Presenta la problemática sobre la base de dos o más ámbitos (política, económica, social o cultural) interrelacionados DIM3 <i>La desigualdad social que ha generado el modelo económico...</i>	Presenta la problemática sobre la base de dos o más ámbitos (política, económica, social o cultural), pero desvinculados entre sí DIM2 <i>Por una parte, ha generado gran pobreza en los adultos mayores. Por otra parte...</i>	Presenta la problemática sobre la base de un único ámbito (política, económica, social o cultural) DIM1 <i>La desigualdad social...</i>	Presenta la problemática desde una dimensión (política, económica, social o cultural) restringida a condiciones coyunturales, personales, cercanas DIM0 <i>A los adultos mayores le escasean los recursos...</i>
4. Actores	Incorpora actores, agrupaciones y/o instituciones relacionadas de forma directa e indirecta en la descripción de la problemática que intervienen y toman decisiones sobre la problemática AC3 <i>El movimiento no más AFP ha decidido... El empresariado propuso...</i>	Incorpora actores, agrupaciones o instituciones relacionadas de forma directa e indirecta con la problemática, pero que actúan pasivamente en ese contexto AC2 <i>El movimiento social se vio afectado...</i>	Solo incorpora actores sociales relacionados de forma directa con la problemática y actuando pasivamente en ella AC1 <i>Los adultos mayores sufren...</i>	Reconoce personajes o sujetos característicos afectados directamente con la problemática AC0 <i>La sra. María...</i>

(CONTINÚA)

Criterios	Nivel 3	Nivel 2	Nivel 1	Nivel 0
5. Causalidad	Explica la problemática a partir de distintas causas (coyunturales y estructurales), considerando dimensiones políticas, económicas, culturales o sociales, que aparecen relacionadas entre sí MUL3 <i>El modelo económico plantea condiciones de inequidad social...</i>	Explica la problemática incorporado causas coyunturales y estructurales en diversas dimensiones, pero sumando elementos, sin relación entre ellos MUL2 <i>En términos sociales... luego, en términos políticos...</i>	Explica la problemática a partir de una enumeración de causas coyunturales, donde prima un tipo de dimensión MUL1 <i>Esta situación generó...</i>	Realiza explicaciones unicasuales, o que confunden causas con consecuentes MULO <i>Porque los extranjeros se niegan...</i>
6. Cadena causal	Organiza de manera lógica y aceptable disciplinariamente causas, motivaciones, hechos y consecuentes CAD3 <i>Esas causas motivan a... lo que impacta en x y genera...</i>	Organiza de manera lógica y aceptable disciplinariamente causas, motivaciones, hechos y consecuentes, aunque uno de los elementos puede ser omitido CAD2 <i>Eso tiene como consecuencia que se motive a... lo que impacta en X y genera...</i>	Evidencia ausencia o confusión de la lógica entre algunos de los elementos de la cadena causal CAD1 <i>Y por esto, él deseó causar...</i>	Los elementos de cadena causal indistinguibles CAD0 <i>Esto no causa...</i>
7. Empatía	Reconoce motivaciones de los actores y agentes sociales implicados al explicar el problema, considerando su contexto histórico EX3 <i>La elite no esperaba...</i>	Reconoce motivaciones de personajes y actores sociales, pero desde lógicas cotidianas personales o atemporales EX2 <i>El ministro pretendía...</i>	Identifica personajes desde estereotipos, víctimas o culpables, sin alusión a comprensión de motivaciones contextualizadas EX1 <i>La culpa es del gobierno porque...</i>	Explica el problema desde la acción o motivación de un personaje en particular. O bien, lo hace desde rasgos de personalidad binarios (bueno/malo, simpático/pesado) EX0 <i>Porque los inquilinos no esperaban democracia...</i>
8. Duración histórica	Al explicar los problemas hacen referencia a continuidades y cambios desde procesos y/o eventos que evidencian corta, media y larga duración DU3 <i>La desigualdad histórica...</i>	Al explicar los problemas hacen referencia a continuidades y/o cambios en procesos y/o eventos de corta, mediana o larga duración DU2 <i>El cambio en el modelo económico...</i>	Solo incorpora marcas textuales de continuidad o cambio relacionadas con procesos y/o eventos coyunturales DU1 <i>Porque el cambio en la constitución...</i>	Plantea explicaciones presentistas, sin referencias históricas DU0 <i>Porque no se quiere modificar...</i>
9. Conceptualización*	Utiliza de manera pertinente conceptos propios de las ciencias sociales CON3 <i>El modelo neoliberal...</i>	Utiliza conceptos relacionados con las ciencias sociales, sin embargo, son confusos en el contexto en que se plantean CON2 <i>El sistema parlamentario...</i>	Utiliza conceptos relacionados con las ciencias sociales, sin embargo, son utilizados con escasa aceptabilidad disciplinar. CON1 <i>El modelo mercantilista...</i>	Utiliza categorías coloquiales para explicar conceptos técnicos CON0 <i>Cuando no hay restricción de tener cosas...</i>

(CONTINUÍA)

Criterios	Nivel 3	Nivel 2	Nivel 1	Nivel 0
10. Significancia histórica	Señala eventos o procesos que confirman sus explicaciones y que son aceptables disciplinariamente SIG3 <i>Tal como ocurrió con la crisis económica de 1930...</i>	Señala eventos o procesos que apoyan sus explicaciones y que son aceptables disciplinariamente, aunque no tienen un rol clave en la explicación SIG2 <i>Cuando había ocurrido la crisis...</i>	Incorpora ideas o eventos con significancia histórica, pero con imprecisiones de aceptabilidad disciplinar. O bien, utiliza referencias de historia contrafactual escasamente pertinentes SIG1 <i>Se dice que si Chile hubiese logrado...</i>	Incorporación de eventos irrelevantes o con errores. O no explicación de eventos o procesos con significancia histórica SIG0 <i>Chile fue el país que perdió más dinero durante la crisis...</i>
11. Interpretación	Identifica la perspectiva de las fuentes de naturaleza inferencial** y analítica,*** considerando cuándo, dónde y/o de quién proviene para cuestionar la fuente (considerar condición de origen) PE3 <i>Considerando que X pertenece a...</i>	Reconoce perspectivas explícitas en fuentes de naturaleza analítica, pero no de naturaleza inferencial, pero considerando su condición de origen PE2 <i>Desde la perspectiva de...</i>	Extrae información textual de la fuente, que podrían apoyar una idea, pero sin considerar su condición de origen en su perspectiva PE1 <i>Como plantea X...</i>	Extrae información textual de la fuente que utiliza en un sentido contradictorio respecto de su origen PE0 <i>Porque el dice que...</i>
12. Forma de la fuente	Utiliza información desde varias fuentes de distinta naturaleza (escrita, iconográfica, etc.) USO3 <i>Como plantea el gráfico y la noticia...</i>	Utiliza varias fuentes, pero todas de la misma naturaleza USO2 <i>La noticia 1 y la 2.</i>	Utiliza una sola fuente en su relato USO1 <i>La noticia...</i>	Utiliza solo una parte de un tipo de fuente (solo los títulos, por ejemplo). O bien, no incorpora otras fuentes USO0 <i>El peso de...</i>
13. Corroboración	Evidencia convergencias y divergencias entre las fuentes COR3 <i>Como señalan X y X a diferencia de Y</i>	Evidencia solo divergencias entre las fuentes o solo convergencias entre ellas COR2 <i>X está en desacuerdo con Y</i>	Presenta fragmentos de información que respaldan puntos de vista, sin reconocer convergencias o divergencias entre ellos COR1 <i>X tanto como X</i>	Presenta fragmentos de información, con escasa relación con el relato elaborado COR0 <i>X</i>
14. Referenciación	Incorpora marcas de autoría de terceros o reconocimientos de otras voces autorales y al hacerlo usa normas de referencia REF3 <i>Valdivia (2009)</i>	Usa marcas de autoría de terceros y reconoce otras voces autorales, pero no utiliza normas de referenciación REF2 <i>Como plantea Valdivia...</i>	Se incorporan marcas de autoría de terceros, pero sin reconocer voces autorales REF1 <i>El autor dice...</i>	El relato carece de marcas de autoría de terceros o se apropia de forma indebida, ya que se pueden advertir marcas de heteroglosia, pero sin referencias a otros autores REF0 <i>He pensado...</i>

(CONTINÚA)

Criterios	Nivel 3	Nivel 2	Nivel 1	Nivel 0
15.Consistencia	Defiende un punto de vista explícito con argumentos fundamentados que persuaden a la audiencia sobre el problema POS3 <i>Es urgente evitar...</i>	Plantea una posición coherente con la evidencia disponible en el relato construido, pero en relación indirecta al problema POS2 <i>Se podría decir que...</i>	Plantea una posición implícita o explícita, pero que tiene escasa relación con los argumentos planteados y con la problemática POS1 <i>Los argumentos son... pero...</i>	La posición del autor es ambigua, ya que no se logra identificar su postura. POS0 <i>En tanto... pero también...</i>
16.Enfoque	La posición y propuesta están fundadas en un enfoque de derechos, equidad social y ética del bien común ENF3 <i>La necesidad de resguardar el derecho a...</i>	La posición y propuesta se plantean sin transgresión evidente en derechos y equidad social generales, pero se plantea de forma muy genérica o desconoce posibles vulneraciones ENF2 <i>Se propone evitar, pero no es posible asegurar que...</i>	Existe un posicionamiento que no trasgrede de forma evidente el enfoque de derechos y equidad social, pero justifica acciones en contra del bien común ENF1 <i>No debería ocurrir, pero se debería pensar en la necesidad de personas como...</i>	La posición y propuesta desconoce el enfoque de derechos y equidad social, o relativiza o justifica su transgresión ENF0 <i>Deberían dejarlos sin trabajo...</i>
17.Propuesta	Genera un propuesta coherente, viable y pertinente para abordar la problemática planteada reconociendo su complejidad PRO3 <i>En ese sentido, se propone avanzar en...</i>	Plantea una propuesta coherente con el problema, pero que resulta poco viable en el contexto PRO2 <i>Se puede acabar con toda desigualdad en el mundo...</i>	Se plantea una propuesta que puede ser viable en el contexto planteado, pero no aborda la problemática de fondo o la simplifica PRO1 <i>Se podría pensar en nuevas tecnologías que...</i>	La propuesta es ingenua o inviable PRO0 <i>Todos deberían sentirse parte de un nuevo mundo...</i>
18.Responsabilidad de acción	Considera la responsabilidad de diversos actores y agentes sociales, con pertinencia a los roles y funciones de las instituciones RES3 <i>La acción comunitaria debería promover que la municipalidad...</i>	Incorpora la responsabilidad de diversos actores y agentes sociales, aunque algunos de ellos tienen poca pertinencia desde su rol o función social Res2 <i>La municipalidad debería cambiar el sistema de pensiones...</i>	La propuesta establecida recae exclusivamente en el rol de un actor o agente social Res1 <i>Las autoridades deberían cambiarlo...</i>	La propuesta establecida recae exclusivamente en el rol de un personaje; o bien, no identifica la responsabilidad de acción Res0 <i>Debería hacerlo la Sra. Bachelet.</i>

* Si el párrafo completo está copiado desde las fuentes, los conceptos no serán evaluados.

** Indican una postura frente a un tema, o bien, señalan un hecho o una manifestación del problema a explicar. Generalmente son fuentes primarias (Henríquez, 2018)

*** Presentan posiciones fundamentadas de el o los enunciadores de modo que permiten sopesar el valor epistémico de su planteamiento. Generalmente refiere a fuentes secundarias (Henríquez, 2018).

Fuente: elaboración propia.

Notas

¹ En adelante, en este artículo se usará el masculino con el único objetivo de hacer más fluida la lectura, sin menoscabo de género.

² Debido a los antecedentes apuntados, se probó el modelo de tres dimensiones,

con la hipótesis de que la dimensión *explicación* contiene la correspondiente a *contextualización* (Lee, 2009). Sin embargo, esta prueba no generó mejores índices de ajuste.

Referencias

- Adams, Raymond (2005). "Reliability as a measurement design effect", *Studies in Educational Evaluation*, vol. 31, núms. 2-3, pp. 162-172.
- Alfageme-Gonzalez, Begoña; Monteagudo, José y Miralles, Pedro (2016). "Concepciones del profesorado de Geografía e Historia sobre la evaluación de los aprendizajes", *Actualidades Investigativas en Educación*, vol. 16, núm. 2, pp. 1-22.
- American Educational Research Association, American Psychological Association y National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*, Washington D.C.: American Educational Research Association.
- Bartelds, Hanneke; Savenije, Geerty y van Boxtel, Carla (2020). "Students' and teachers' beliefs about historical empathy in secondary history education", *Theory and Research in Social Education*, vol. 48 núm 4, pp. 529-551. <https://doi.org/10.1080/00933104.2020.1808131>
- Barton, Keith (2005). "Best not to forget them": Secondary students' judgments of historical significance in Northern Ireland", *Theory & Research in Social Education*, vol. 33, núm. 1, pp. 9-44. <https://doi.org/10.1080/00933104.2005.10473270>
- Breakstone, Joel; Smith, Mark y Wineburg, Sam (2013). "Beyond the bubble in history/social studies assessments", *Phi Delta Kappan*, vol. 94, núm. 5, pp. 53-57. <https://doi.org/10.1177/003172171309400512>
- Cervini, Rubén Alberto; Dari, Nora y Quiroz, Silvia (2015). "Género y rendimiento escolar en América Latina. Los datos del SERCE en matemática y lectura", *Revista Iberoamericana de Educación*, vol. 68, pp. 99-116. <https://doi.org/10.35362/rie680206>.
- College Board (2017). *AP United States history: 2017 free-response questions*, Nueva York: College Board.
- Contreras-Salinas, Sylvia; Olavarría Aranguren, José; Celedón Bulnes, Roberto y Molina Gutiérrez, Rodrigo (2020). "Factores asociados a la brecha de género en lectura entre estudiantes de establecimientos de enseñanza secundaria en Chile", *Perfiles Educativos*, vol. 42, núm. 170, pp. 60-76. <https://doi.org/10.22201/iisue.24486167e.2020.170.59013>
- De La Paz, Susana; Ferretti, Ralph; Wissinger, Daniel; Yee, Laura y MacArthur, Charles (2012). "Adolescents' disciplinary use of evidence, argumentative strategies, and organizational structure in writing about historical controversies", *Written Communication*, vol. 29, núm. 4, pp. 412-454. <https://doi.org/10.1177/0741088312461591>
- De La Paz, Susana y Wissinger, Daniel (2015). "Effects of genre and content knowledge on historical thinking with academically diverse high school students", *Journal of Experimental Education*, vol. 83, núm. 1, pp. 110-129. <https://doi.org/10.1080/00220973.2013.876228>

- De La Paz, Susana; Monte-Sano, Chauncey; Felton, Mark; Croninger, Robert; Jackson, Cara y Piantedosi, Kelly Worland (2017). "A historical writing apprenticeship for adolescents: Integrating disciplinary learning with cognitive strategies", *Reading Research Quarterly*, vol. 52, núm. 1, pp. 31-52. <http://dx.doi.org/10.1002/rrq.147>
- Embretson, Susan (ed.) (2010). *Measuring psychological constructs: Advances in model-based approaches*, Washington D.C.: American Psychological Association.
- Ercikan, Kadriye y Seixas, Peter (2015). *New directions in assessing Historical Thinking*, Nueva York: Routledge.
- Evans, Ronald W. y Saxe, David Warren (eds.) (1996). *Handbook on teaching social issues*, Washington D.C.: National Council for Social Studies.
- Fisher, William y Wilson, Mark (2019). "The BEAR Assessment System Software as a platform for developing and applying UN SDG metrics", *Journal of Physics: Conference Series*, núm. 1379. <https://doi.org/10.1088/1742-6596/1379/1/012041>
- Gerwin, David (2003). "Responding to a social studies classic—"Decision making: The heart of social studies instruction", *The Social Studies*, vol. 94, núm. 1, pp. 23-27. <https://doi.org/10.1080/00220973.1945.11019962>
- Grant, S. G. (2017). "The problem of knowing what students know: Classroom-based and large-scale assessment in social studies", en M. McGlinn y C. Mason (eds.), *The Wiley Handbook of Social Studies Research*, Londres: John Wiley y Sons, pp. 461-476.
- Grant, S. G.; Gradwell, Jill y Cimbricz, Sandra (2004). "A question of authenticity: The document-based question as an assessment of students' knowledge of history", *Journal of Curriculum and Supervision*, núm. 19, pp. 309-337.
- Gutiérrez, Martha y Pagès, Joan (2018). *Pensar para intervenir en la solución de las injusticias sociales*, col. Trabajos de Investigación núm. 1, Pereira: Universidad Tecnológica de Pereira.
- Hahn, Carole (2020). "Educating citizens in an age of globalization, migration, and transnationalism: A study in four European democracies", *Theory and Research in Social Education*, vol. 42, núm. 2, pp. 244-184. <https://doi.org/10.1080/00933104.2019.1707139>
- Henríquez, Rodrigo; Carmona, Andrés; Quinteros, Alen y Garrido, Mabelin (2018). *Leer y escribir para aprender Historia: Secuencias para la enseñanza y el aprendizaje del pensamiento histórico*, col. Educación. núm. 1, Santiago: Ediciones UC.
- Hooper, Daire; Coughlan, Joseph y Mullen, Michael (2008). "Structural Equation Modelling: Guidelines for Determining Model Fit", *Electronic Journal of Business Research Methods*, vol. 6, núm. 1, septiembre, pp. 53-60.
- Huijgen, Tim; Van Boxtel, Carla; Van de Grift, Wim y Holthuis, Paul (2017). "Toward historical perspective taking: students' reasoning when contextualizing the actions of people in the past", *Theory y Research in Social Education*, vol. 45, núm. 1, pp. 110-144. <https://doi.org/10.1080/00933104.2016.1208597>
- Huijgen, Tim; Van de Grift, Wim; Van Boxtel, Carla y Holthuis, Paul (2018). "Promoting historical contextualization: The development and testing of a pedagogy", *Journal of Curriculum Studies*, vol. 50, núm. 3, pp. 410-434. <https://doi.org/10.1080/00220272.2018.1435724>

- Huijgen, Tim; Holthuis, Paul; Van Boxtel, Carla; Van De Grift, Wim; y Suhre, Cor (2019). "Students' historical contextualization and the cold war", *British Journal of Educational Studies*, vol. 67 núm 4, pp. 439-468. <https://doi.org/10.1080/00071005.2018.1518512>
- Kane, Michael (2020). "Validity studies commentary", *Educational Assessment*, vol. 25, núm. 1, pp. 83-89. <https://doi.org/10.1080/10627197.2019.1702465>
- Kemp, Robin (2011). "Thematic or sequential analysis in causal explanations? Investigating the kinds of historical understanding that Year 8 and Year 10 demonstrate in their efforts to construct narratives", *Teaching History*, vol 1, núm. 145, pp. 32-43.
- Larroucau, Tomas; Ríos, Ignacio y Mizala, Alejandra (2015). "Efecto de la incorporación del ranking de notas en el proceso de admisión a las universidades chilenas", *Pensamiento Educativo*, vol. 52, núm. 1, pp. 95-118. <https://doi.org/10.7764/PEL.52.1.2015.8>
- Lee, Peter (2009). "Is any explanation better than none?", *Teaching History*, vol. 1, núm. 137, pp. 42-49.
- Levstik, Linda y Tyson, Cynthia (eds.) (2008). *Handbook of Research in Social Studies Education*, Nueva York/Londres: Routledge.
- Levy, Brett (2018). "Youth developing political efficacy through social learning experiences: Becoming active participants in a supportive Model United Nations Club", *Theory and Research in Social Education*, vol. 46, núm. 3, pp. 410-448. <https://doi.org/10.1080/00933104.2017.1377654>
- McGlinn, Meghan y Mason, Cheryl (eds.) (2017). *The Wiley Handbook of Social Studies Research*, Londres: John Wiley y Sons, Inc.
- McGrew, Sarah; Breakstone, Joel; Ortega, Teresa; Smith, Mark y Wineburg, Sam (2018). "Can students evaluate online sources? Learning from assessments of civic online reasoning", *Theory and Research in Social Education*, vol. 46, núm. 2, pp. 165-193. <https://doi.org/10.1080/00933104.2017.1416320>
- Miralles, Pedro; Gómez, Cosme y Monteagudo, José (2012). "La evaluación de la competencia social y ciudadana en ciencias sociales al finalizar las etapas de primaria y secundaria", *Investigación en la Escuela*, núm. 78, pp. 19-30.
- Monte-Sano, Chauncey y De la Paz, Susan (2012). "Using writing tasks to elicit adolescents' historical reasoning", *Journal of Literacy Research*, vol. 44, núm. 3, pp. 273-299. <https://doi.org/10.1177/1086296X12450445>
- Monte-Sano, Chauncey y Quince, Christine (2021). "Reflections on designing curriculum to interrogate social studies", *Journal of Adolescent and Adult Literacy*, vol. 64, núm. 5, pp. 575-580. <https://doi.org/10.1002/jaal.1137>
- Murillo, Johanna y Martínez, Carlos (2014). "Habilidades de pensamiento social: describir, explicar, interpretar y argumentar en el aula", *Itinerario Educativo*, vol. 28, núm. 64, pp. 103-125. <https://doi.org/10.21500/01212753.1421>
- National Council for the Social Studies (ed.) (2013). *College, Career y Civic Life (C3) Framework for social studies state standards: Guidance for enhancing the rigor of K-12 Civics, Economics, Geography and History*, Silver Spring: National Council for the Social Studies.
- Niepel, Christoph; Mustafić, Maida; Greiff, Samuel y Roberts, Richard D. (2015). "The dark side of creativity revisited: Is students' creativity associated with subsequent

- decreases in their ethical decision making?”, *Thinking Skills and Creativity*, vol. 18, diciembre, pp. 43-52. <https://doi.org/10.1016/j.tsc.2015.04.005>
- Ortuño, Jorge; Ponce, Ana y Serrano, Francisca (2016). “La idea de causalidad en las explicaciones históricas del alumnado de educación primaria”, *Revista de Educación*, núm. 371, enero-marzo, pp. 9-34.
- Pagés, Joan (2002). “La formación del pensamiento social”, en Benejam, Pilar y Pagés, Joan (coords.), *Enseñar y aprender ciencias sociales, geografía e historia en la educación secundaria*, núm. 1, Barcelona: Horsi/Instituto de Ciencias de la Educación-Universidad de Barcelona, pp. 151-168.
- Pagés, Joan y Santisteban, Antoni (coords.) (2011). *Les qüestions socialment vives i l'ensenyament de les ciències socials*, Barcelona: Universitat Autònoma de Barcelona.
- Pearcy, Mark (ed.) (2017). *Best practices in social studies assessment*, Charlotte: Information Age Publishing.
- Pedrozo, Sueila (2011). “To be “cool” or not to be “cool”: Young people’s insights on consumption and social issues in Rio de Janeiro”, *Journal of Youth Studies*, vol. 14, núm. 1, pp. 109-123. <https://doi.org/10.1080/13676261.2010.506525>
- Pipkin, Diana y Sofía, Paula (2005). “La formación del pensamiento social en la escuela media: factores que facilitan y obstaculizan su enseñanza”, *Clio & Asociados: la Historia Enseñada*, núm. 8, pp. 85-94. <https://doi.org/10.14409/cya.v1i8.1593>
- Raykov, Tenko (1998). “Coefficient alpha and composite reliability with interrelated nonhomogeneous items”, *Applied Psychological Measurement*, vol. 22, núm. 4, pp. 375-385. <https://doi.org/10.1177/014662169802200407>
- Reagan, Rebecca (2008). “Direct instruction in skillful thinking in fifth-grade American History”, *The Social Studies*, vol. 99, núm. 5, pp. 217-222. <https://doi.org/10.3200/TSS.99.5.217-222>
- Reich, Gabriel (2011). “Testing collective memory: Representing the Soviet Union on multiple-choice questions”, *Journal of Curriculum Studies*, vol. 43, núm. 4, pp. 507-532. <https://doi.org/10.1080/00220272.2011.578665>
- Reisman, Abby (2012). “The ‘Document-Based Lesson’: Bringing disciplinary inquiry into high school history classrooms with adolescent struggling readers”, *Journal of Curriculum Studies*, vol. 44, núm. 2, pp. 233-264. <https://doi.org/10.1080/00220272.2011.591436>
- Reisman, Avishag, y Wineburg, Sam (2008). “Teaching the skill of contextualizing in History”, *The Social Studies*, vol. 99, núm. 5, pp. 202-207. <https://doi.org/10.3200/TSS.99.5.202-207>
- Reisman, Abby; Brimesk, Emily y Hollywood, Claire (2019). “Assessment of Historical Analysis and Argumentation (AHAA): A new measure of document-based historical thinking”, *Cognition and Instruction*, vol. 37, núm. 4, pp. 534-561. <https://doi.org/10.1080/07370008.2019.1632861>
- Ross, E. Wayne (2017). *Rethinking of social studies: Critical pedagogy in pursuit of dangerous citizenship*, Charlotte: Information Age Publishing.
- Ross, E. Wayne (2019). “The challenges of teaching social studies: What teachers? What citizenship? What future? A Tribute to Joan Pagès Blanch”, en M. Ballbé, N.

- González-Monfort y A. Santisteban, A. (eds.), *Quin professorat, quina ciutadania, quin futur? Els reptes de l'ensenyament de les ciències socials, la geografia i la història*, Barcelona: Universitat Autònoma de Barcelona-GREDICS/Ajuntament de Barcelona, pp. 39-52.
- Saye, John (2017). "Disciplined inquiry in social studies classrooms", en M. McGlinn y C. Mason (eds.), *The Wiley Handbook of Social Studies Research*, Londres: John Wiley y Sons, pp. 336-359.
- Schreiber, James; Nora, Amaury; Stage, Frances K.; Barlow, Elizabeth A. y King, Jamie (2006). "Reporting structural equation modeling and confirmatory factor analysis results: A review", *The Journal of Educational Research*, vol. 99, núm. 6, pp. 323-338. <https://doi.org/10.3200/JOER.99.6.323-338>
- Seixas, Peter (2017). "A model of historical thinking", *Educational Philosophy and Theory*, vol. 49, núm. 6, pp. 593-605. <https://doi.org/10.1080/00131857.2015.1101363>
- Siegfried, John; Krueger, Alan; Collins, Susan; Frank, Robert; MacDonald, Richard; McGoldrick, Kim, Marie; Taylor, John y Vredevel, George (2010). *Voluntary National Content Standards in Economics*, Nueva York: Council for Economic Education.
- Smith, Dan (2014). "Period, place and mental space: using historical scholarship to develop year 7 pupils' sense of period", *Teaching History*, vol. 1, núm. 154, marzo, pp. 8-16.
- Smith, Mark (2017). "Cognitive validity: Can multiple-choice items tap historical thinking processes?", *American Educational Research Journal*, vol. 54, núm. 6, pp. 1256-1287. <https://doi.org/10.3102/0002831217717949>
- Smith, Mark (2018). "New multiple-choice measures of historical thinking: An investigation of cognitive validity", *Theory & Research in Social Education*, vol. 46, núm. 1, pp. 1-34. <https://doi.org/10.1080/00933104.2017.1351412>
- Smith, Mark; Breakstone, Joel y Wineburg, Sam (2018). "History assessments of thinking: A validity study", *Cognition and Instruction*, vol. 1, núm. 27, pp. 118-144. <https://doi.org/10.1080/07370008.2018.1499646>
- Stoel, Gerhard; Van Drie, Jannet y Van Boxtel, Carla (2015). "Teaching towards historical expertise. Developing a pedagogy for fostering causal reasoning in history", *Journal of Curriculum Studies*, vol. 47, núm. 1, pp. 49-76. <https://doi.org/10.1080/00220272.2014.968212>
- Supianto, Kumaidi y Suryono, Yoyon (2021). "Development of project-based historical knowledge assessment instruments", *Historical Encounters*, vol. 8, núm. 1, pp. 16-29. <https://doi.org/10.52289/hej8.102>
- Torrez, Cheryl y Claunch-Lebsack, Elisabeth (2013). "Research on assessment in the social studies classroom", en J. McMillan (ed.), *Handbook Research on Classroom Assessment*, Thousand Oaks: SAGE Publications, pp. 461-472. <https://dx.doi.org/10.4135/9781452218649.n26>
- Torrez, Cheryl y Claunch-Lebsack, Elisabeth (2014). "The present absence: Assessment in social studies classrooms", *Action in Teacher Education*, vol. 36, núms. 5-6, pp. 559-570. <https://doi.org/10.1080/01626620.2014.977756>
- Unesco (ed.) (2016). *Challenging inequalities: Pathways to a just world*, París: Unesco Publishing.

- Van Drie, Jannet y Van Boxtel, Carla (2008). "Historical reasoning: Towards a framework for analyzing students' reasoning about the past", *Educational Psychology Review*, vol. 20, núm. 2, pp. 87-110. <https://doi.org/10.1007/s10648-007-9056-1>
- VanSledright, Bruce y Limón, Margarita (2006). "Learning and teaching social studies: A review of cognitive research in History and Geography", en P. Alexander y P. Winne (eds.), *Handbook of Educational Psychology*, núm. 5, Nueva York/Londres: Routledge, pp. 545-570.
- Wainer, Howard (2011). *Uneducated Guesses*, Nueva Jersey: Princeton University Press.
- Wilson, Mark (2005). *Constructing measuring: An item response modeling approach*, Nueva Jersey: Lawrence Erlbaum Associates.

Artículo recibido: 30 de junio de 2021

Dictaminado: 7 de marzo de 2022

Segunda versión: 19 de abril de 2022

Aceptado: 9 de mayo de 2022