



Original

Efectos de un programa de formación del profesorado en la motivación y satisfacción de los estudiantes de historia en enseñanza secundaria

Cosme J. Gómez Carrasco^a, Jairo Rodríguez-Medina^{b,*},
Pedro Miralles Martínez^a, y Víctor Benito Arias González^c

^a Universidad de Murcia, Murcia, España

^b Universidad de Valladolid, Valladolid, España

^c Universidad de Salamanca, Salamanca, España

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 18 de mayo de 2020

Aceptado el 23 de julio de 2020

On-line el 28 de agosto de 2020

Palabras clave:

Educación histórica

Metodología didáctica

Formación del Profesorado

Motivación

Satisfacción

R E S U M E N

Las clases de historia en Educación Secundaria forman un espacio de enseñanza y aprendizaje con unas prácticas que se nutren de dos modelos epistemológicos y metodológicos tradicionales. Frente a estos modelos, se realiza un programa formativo en la especialidad de Geografía e Historia del Máster de Formación del Profesorado destinado a cambiar sus concepciones metodológicas y epistemológicas y al diseño de unidades formativas de innovación. Se evalúan sus efectos a través de la implementación de estas unidades formativas en la fase de prácticas curriculares. Esta evaluación se realiza con un pretest y postest formado por dos escalas (motivación y satisfacción). La estrategia de análisis se desarrolla en dos fases. En la primera se comprueba la invarianza factorial longitudinal, analizando progresivamente dicha invarianza entre las escalas pretest y postest. En la segunda fase, se aplica un modelo de crecimiento de segundo orden o curva de factores para evaluar el cambio en las variables latentes entre el pretest y el postest. Tras la aplicación del programa, se observa un aumento sustancial en la motivación y satisfacción del alumnado. Este incremento es más visible en motivación, debido a los bajos niveles de los que parten los estudiantes.

© 2020 Universidad de País Vasco. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Effects of a teacher training program on the motivation and satisfaction of History secondary students

A B S T R A C T

The history classes in Secondary Education form a teaching and learning space with practices that are nourished by two traditional epistemological and methodological models. Faced with these models, a training program was carried out in the specialty of Geography and History of the Master's Degree in Teacher Training. This program was aimed at changing its methodological and epistemological conceptions and the design of innovative training units. Its effects have been evaluated through the implementation of these training units in the curricular practical phase. This evaluation has been carried out with a pretest and posttest consisting of two scales (motivation and satisfaction). The analysis strategy was developed in two phases. In the first, the longitudinal factor invariance was verified, progressively analyzing this invariance between the pretest and posttest scales. In the second phase, a second-order growth model

Keywords:

History education

Didactic methodology

Teacher Education

Motivation

Satisfaction

* Autor para correspondencia.

Correo electrónico: jairo.rodriguez.medina@uva.es (J. Rodríguez-Medina).

or factor curve was applied to assess the change in latent variables between the pretest and posttest. After the application of the program, a substantial increase in student motivation and satisfaction was observed. This increase was more visible in motivation, due to the low levels from which the students start.

© 2020 Universidad de País Vasco. Published by Elsevier España, S.L.U. All rights reserved.

Introducción

Tradicionalmente las clases de historia en Educación Secundaria forman un espacio de enseñanza y aprendizaje con unas prácticas asumidas generalmente por sus miembros. El alumnado con más éxito escolar adopta las formas esperadas por esta comunidad para leer, memorizar, pensar y escribir como les exigen (Nokes, 2017). Las actividades planteadas en las clases de historia, la interacción habitual con el profesorado, y los procedimientos y técnicas de evaluación planteados por el docente refuerzan la reproducción de estas prácticas (Gómez et al., 2020). Estas prácticas se nutren de dos modelos, uno epistemológico y otro metodológico. El primero es el modelo de Historias Generales, un enfoque de enseñanza de la historia perfilado en el siglo XIX, desde el que se pretende mostrar la totalidad de los hechos históricos de forma cronológica (Valls y Colomer, 2018). El segundo de los modelos procede de una tradición metodológica que entiende la práctica de enseñanza como mera reproducción del conocimiento construido fuera de las aulas (Merchán, 2011).

Desde hace varias décadas, los trabajos desde el área de educación histórica han intentado reinventar esos modelos, aunque haciendo más énfasis en la transformación de las concepciones epistemológicas que en las prácticas educativas en el aula. Así, los estudios de Monte-Sano (2011), Reisman (2012), Van Boxtel y van Drie (2012), están relacionados con temas como pensamiento histórico, alfabetización histórica o el uso de fuentes primarias en el aula. Desde otro enfoque, los trabajos de Carretero y van Alphen (2014), Grever et al. (2011) y López et al. (2014) se han centrado en temáticas de conciencia histórica, identidad y memoria histórica.

Estos estudios pretenden reinventar el papel del alumnado en la clase de historia: que los estudiantes entiendan que aprender esta disciplina no es simplemente retener en la memoria una narración canonizada (Monte-Sano et al., 2014). Aprender historia supone un trabajo cognitivo diferente que permite construir interpretaciones del pasado basadas en fuentes y pruebas (VanSledright, 2014). En contraste a lo que se hace habitualmente en el aula, donde los textos se usan para transmitir información a los estudiantes, que suelen memorizarlos, los historiadores ven los textos como pruebas para construir explicaciones sobre los procesos históricos (Lesh, 2011).

Para mejorar la educación histórica es necesario que los docentes apuesten por metodologías alternativas de enseñanza, pero estas deben ir acompañadas de un cambio epistemológico. Tanto los docentes como el alumnado deben cambiar los lastres epistemológicos que conciben la historia como un conjunto de conocimientos cerrados. Por ello es tan importante una intervención en la formación del profesorado que tenga como objetivos la mejora de los conocimientos, habilidades y competencias docentes relacionados con estos ámbitos metodológicos y epistemológicos (Barnes et al., 2017).

El presente estudio

En los últimos años se han incrementado los análisis sobre métodos y estrategias en la clase de historia, principalmente sobre el uso de recursos digitales (Miralles et al., 2019). En España las investigaciones empíricas en el área de educación histórica se han centrado principalmente en análisis de diagnóstico, sobre todo con libros

de texto, exámenes y concepciones del profesorado y alumnado (Cuenca et al., 2017). No son abundantes, sin embargo, investigaciones evaluativas.

El objetivo de este trabajo es analizar los efectos de un programa de formación del profesorado sobre la motivación y satisfacción de los estudiantes de Educación Secundaria (Maloy et al., 2019). Para alcanzar este objetivo se plantean las siguientes preguntas: (1) ¿Cómo cambia la motivación del alumnado con la implementación de las unidades formativas?; y (2) ¿Cómo cambia la satisfacción del alumnado con la implementación de las unidades formativas?

Método

Participantes

Participan 473 alumnos de Educación Secundaria de 18 aulas de la Región de Murcia (52% mujeres). Estas 18 aulas están situadas en 14 centros, 13 de carácter público y uno privado. Las unidades didácticas se implementan en los cuatro cursos de la ESO y en los dos cursos de Bachillerato. La elección de los participantes de la muestra por conveniencia está relacionada con la asignación de los centros de prácticas curriculares a los docentes en formación que diseñan e implementan las unidades formativas.

Instrumentos

Se diseña un cuestionario compuesto por dos subescalas (*motivación y satisfacción*) de acuerdo con otros estudios que han evaluado programas de intervención basados en métodos activos (Da Rocha et al., 2016). La validación del contenido se realiza mediante el procedimiento *interjueces* en torno a las categorías de pertinencia y claridad de los ítems del instrumento, y a través de un grupo de discusión con siete expertos. Se opta por el método Delphi y, tras las modificaciones pertinentes, se realiza una segunda ronda con los expertos para validar definitivamente los dos instrumentos, divididos en dos subescalas. Se obtienen buenos resultados de concordancia utilizando el coeficiente de acuerdo ponderado de Bangdiwala (Bangdiwala, 1987). Específicamente, .86 para la claridad de los ítems y .91 para la relevancia de los ítems.

La primera subescala, relativa a la motivación de los estudiantes, está compuesta por seis ítems agrupados en *motivación intrínseca* («Las clases me motivan a conocer más sobre la historia») y *motivación extrínseca* («Las clases de historia me motivan para conseguir mejores notas»). La segunda subescala se refiere a la *satisfacción* de los estudiantes y contiene seis ítems. Ejemplo de estos ítems: «Estoy satisfecho/a con el papel que he tenido como alumnado» o «Estoy satisfecho con la forma en la que el profesor ha tratado el tema». El cuestionario también incluye información sobre las características del alumnado: edad, sexo y curso. Se establece una escala likert del 1-5, entre (1) «muy en desacuerdo» y (5) «muy de acuerdo».

Se comprueba la consistencia interna de las dos subescalas y de la puntuación total mediante los coeficientes α de Cronbach para datos ordinales y ω de McDonald (McDonald, 2013). En conjunto, se obtienen valores α ordinal de .92 en el pretest y .92 en el posttest, y coeficientes de fiabilidad global compuesta de .91 en la prueba previa y .92 en la prueba posterior (Rodríguez-Medina et al., 2020). Ambos valores se consideran excelentes. En la escala de motivación de los estudiantes, se obtienen índices de fiabili-

dad que pueden considerarse adecuados (pretest $\alpha = .89$, $\omega = .90$; posttest $\alpha = .89$, $\omega = .92$). Del mismo modo, se obtienen índices de consistencia interna adecuados para la escala de satisfacción de los estudiantes (pretest $\alpha = .82$, $\omega = .85$; posttest $\alpha = .82$, $\omega = .85$).

Procedimiento

Se diseña un programa de intervención en la asignatura «Métodos y recursos para la enseñanza de Geografía, Historia e Historia del Arte» de la especialidad de Geografía e Historia del Máster de Formación del Profesorado (Rodríguez-Medina et al., 2020). Se plantea como objetivo principal mejorar las competencias de los futuros docentes en el diseño de actividades y unidades formativas. El programa formativo consta de ocho sesiones de cuatro horas cada una. La primera sesión se dedica a responder a la pregunta «¿Por qué es necesario un cambio de modelo didáctico para la enseñanza de la Geografía e Historia?» y se une con el trabajo previo que desarrollan sobre competencias históricas. En la segunda y tercera sesión se ejemplifican esos métodos de indagación. Las dos siguientes sesiones se dedican al trabajo con fuentes primarias, patrimonio y recursos digitales. Las tres últimas sesiones se dedican al diseño de las unidades didácticas, aplicando el trabajo teórico a la unidad formativa que se implementa.

Para evaluar la fidelidad de la intervención se diseña una lista de comprobación que contiene las doce estrategias y técnicas que se trabajan principalmente en el programa de intervención (ver documento suplementario 1 en el material adicional). Seis de esas estrategias son de carácter metodológico. Las otras seis técnicas y estrategias son de carácter epistemológico, relacionadas con las competencias históricas propuestas por Seixas y Morton (2013). Respecto a las variables metodológicas, en la mayoría de las unidades didácticas se utilizan recursos digitales, coloquios y debates, portafolios y trabajos de clase. Las técnicas cooperativas se utilizan en dos tercios de las unidades formativas, mientras que las actividades de indagación se usan en la mitad de las intervenciones. La observación directa solo se utiliza en cuatro de cada diez casos (Tabla 1). En el caso de las variables epistemológicas, la mayoría de las intervenciones usan trabajo con fuentes y actividades de causas y consecuencias, y de cambios y permanencias. Las actividades sobre la dimensión ética se utilizan en algo más de la mitad de las unidades formativas. En cambio, las actividades de perspectiva histórica y de relevancia histórica tienen menos presencia.

Para la recogida de datos se opta por dos ocasiones de medición: la primera semana de las prácticas curriculares y al finalizar la unidad formativa. Se obtiene el permiso del comité ético de la Universidad de Murcia para el procedimiento de la investigación (ver documento suplementario 2 en el material adicional). Igualmente se cuenta con el consentimiento informado del alumnado de Educación Secundaria participante (ver documento suplementario 3 en el material adicional).

Análisis de datos

Se calculan las medias y desviaciones típicas de los ítems en el pretest y el posttest y se aplica la prueba de los rangos con signo de Wilcoxon para comprobar si existen diferencias en la percepción de los estudiantes. Para calcular el tamaño del efecto se emplea el estadístico r ($r = Z/\sqrt{N}$; $r > .10$ tamaño del efecto pequeño, $r > .30$ tamaño del efecto medio y $r > .50$ tamaño del efecto grande). A continuación, la estrategia de análisis se desarrolla en dos fases. En la primera de ellas, tras comprobar la adecuación de los datos para el análisis factorial, mediante la prueba de esfericidad de Barlett ($\chi^2 = 7799.53$, $p < .001$) y el índice de adecuación muestral de Kaiser-Meyer-Olkin ($KMO = .91$), se comprueba la invarianza factorial longitudinal. Para ello se analiza progresivamente la invarianza factorial entre las escalas pretest y posttest. Se sigue el procedimiento propuesto

por Liu et al. (2017) para la comprobación de la invarianza longitudinal con datos ordinales, así como para evaluar la relevancia práctica de la violación de la invarianza mediante análisis de sensibilidad.

Se compara el ajuste de las diferentes estructuras obtenidas mediante análisis factorial confirmatorio. Todos los modelos se estiman mediante mínimos cuadrados ponderados «WLSMV». Se contrasta la bondad de ajuste mediante los índices de ajuste comparativo (CFI) y Tucker-Lewis (TLI) y la raíz del error cuadrático medio de aproximación (RMSEA). Se considera que índices CFI y TLI superiores a .90 indican grados de ajuste aceptables y por encima de .95 buenos (Hu y Bentler, 1999). En el caso del RMSEA, valores iguales o inferiores a .05 se interpretan como buenos e inferiores a .08 como aceptables (Browne y Cudeck, 1992; Hu y Bentler, 1999). Se siguen las recomendaciones de Chen (2007) y Cheung y Rensvold (2002), según las cuales incrementos menores a .010 en CFI y TLI y decrementos menores a .015 en RMSEA sugieren que no hay cambios relevantes en el ajuste de un modelo respecto del siguiente más restrictivo, para establecer la relevancia de las diferencias de ajuste entre modelos.

En la segunda fase, se aplica un modelo de crecimiento de segundo orden o curva de crecimiento de los factores (McArdle, 1988) para evaluar el cambio en las variables latentes entre el pretest y el posttest. Este modelo combina un modelo de medición y un modelo de crecimiento en una sola especificación, de tal forma que integra ambos en una sola estructura subyacente. Todos los análisis se realizan utilizando el paquete *lavaan* (Rosseel, 2012) en el software estadístico gratuito R 3.6.3 (R Core y Team., 2020) para el modelo de crecimiento de segundo orden y Mplus 7.0 (Muthén y Muthén, 2015) para el análisis de invarianza.

Resultados

En la Tabla 2 se presentan las medias en el pretest y el posttest, las diferencias entre medias y los resultados de las pruebas de los rangos con signo de Wilcoxon para cada una de las variables observadas en las dos ocasiones de medición. Como se puede comprobar, estos estadísticos indican un incremento significativo en la valoración de cada ítem de ambas subescalas en la prueba realizada tras la intervención.

En las variables *motivación* ($Z = 9.4$, $p < .001$, $r = .443$) y *satisfacción* ($Z = 8.68$, $p < .001$, $r = .408$) el tamaño medio del efecto es entre moderado y alto. En el caso de la subescala de motivación, destaca la valoración que realiza el alumnado sobre la mejora de su motivación para aprender y esforzarse, y su motivación para aprender más sobre la historia. En la subescala de *satisfacción* hay que resaltar la valoración que el alumnado realiza del papel del profesorado y el clima de trabajo en el aula.

Invarianza longitudinal

En la Tabla 3 se presentan los índices de ajuste para cada nivel de invarianza factorial longitudinal en la subescala de motivación. La hipótesis de invarianza configural es aceptable, ya que los seis ítems cargan positivamente en un solo factor en cada ocasión de medición. Los índices muestran un ajuste adecuado del modelo. No se observa ningún cambio en el índice de ajuste comparativo CFI ($\Delta CFI = 0$) respecto al modelo de invarianza configural. El modelo de invarianza fuerte de los datos produce también una $\Delta\chi^2$ significativa ($p < .05$) entre la invarianza métrica y escalar, pero solo un ligero cambio en el CFI ($\Delta CFI < .01$). Finalmente, se examina la invarianza estricta. El modelo de invarianza estricta incrementa la prueba de ajuste χ^2 ; como era de esperar, los resultados mues-

Tabla 1
Variables metodológicas y epistemológicas introducidas en las unidades didácticas

	Número	Porcentaje
<i>Variables metodológicas</i>		
Recursos digitales	17	94.44
Coloquios y debates	17	94.44
Uso de portfolio y trabajos de clase	17	94.44
Técnicas cooperativas	12	66.67
Actividades de indagación	9	50.00
Observación directa, rúbricas o escalas de observación	7	38.89
<i>Variables epistemológicas</i>		
Actividades de cambios/permanencias	16	88.89
Uso de documentos históricos	15	83.33
Actividades de causas/consecuencias	13	72.22
Actividades de dimensión ética	10	55.56
Actividades de perspectiva histórica	3	16.67
Actividades de relevancia histórica	3	16.67

Tabla 2
Diferencias entre las puntuaciones pretest-postest en cada escala

Ítems	Pretest M	Postest M	Dif.	Z	p	Tamaño del efecto r
Motivación						
La forma de presentar y trabajar el tema me ha motivado para conocer más sobre la historia	3.11	3.92	.81	10.12	<.001	.475
He mejorado mi motivación para aprender y esforzarme más en clase	2.98	3.81	.83	10.94	<.001	.516
Mi motivación ha mejorado porque comprendo mejor la realidad social y cultural con la que estoy en contacto	3.29	3.85	.56	7.54	<.001	.358
El tema ha mejorado mi motivación para conseguir mejores notas	3.21	3.78	.57	7.78	<.001	.365
Me he motivado porque podía aportar mi punto de vista o conocimientos propios	2.82	3.68	.86	10.01	<.001	.471
Me he motivado porque hemos utilizado recursos diferentes al libro de texto (Internet, audiovisuales, documentos históricos, etc.)	3.21	4.12	.91	10.01	<.001	.471
Satisfacción						
Estoy satisfecho/a con el papel que he tenido como alumno/a	3.51	4.05	0.54	8.68	<.001	.408
Estoy satisfecho/a con clima de trabajo generado en el aula	3.44	4.06	0.62	9.17	<.001	.431
Estoy satisfecho/a con el trabajo de mis compañeros/as en pequeños grupos	3.25	3.91	0.66	8.98	<.001	.422
Estoy satisfecho/a con los aprendizajes conseguidos	3.60	4.20	0.6	9.05	<.001	.426
Estoy satisfecho/a con la forma en la que el profesor ha tratado el tema	3.55	4.32	0.77	9.47	<.001	.445
Estoy satisfecho/a con mi calificación	3.41	3.93	0.52	6.74	<.001	.316

Tabla 3
Análisis de la invarianza factorial longitudinal de la variable motivación

Motivación	Invarianza configural	Invarianza métrica	Invarianza fuerte	Invarianza estricta
χ^2 / gl	82.860 / 47	91.470 / 52	167.806 / 69	214.664 / 75
$p \chi^2$.001	0.001	0	0
$\Delta\chi^2 / \Delta gl$	-	9 / 5 *	76 / 17 ***	47 / 6 **
$p(\Delta\chi^2)$	-	.028	<.001	<.001
CFI	.998	.998	.995	.993
ΔCFI	-	0	.003	.002
TLI	.997	.997	.995	.993
RMSEA (IC 90%)	.040 (.026 - .054)	.040 (.026 - .054)	.055 (.045 - .066)	.063 (.053 - .073)
p RMSEA $\leq .05$.863	.878	.196	.015
SRMR	.045	.046	.046	.048

CFI: índice de ajuste comparativo, gl : grados de libertad, IC: intervalo de confianza, RMSEA: error cuadrático medio de aproximación, SRMR: raíz del residuo estandarizado medio, TLI: Tucker-Lewis index, ΔCFI : diferencia entre el índice de ajuste comparativo del modelo y el anterior menos restrictivo, Δgl : diferencia entre los grados de libertad de un modelo y el anterior menos restrictivo, $\Delta\chi^2$: diferencia del estadístico χ^2 entre un modelo y el anterior menos restrictivo..

Un asterisco (*) junto a $\Delta\chi^2$ indica que la diferencia con el ajuste del modelo especificado previamente es estadísticamente significativa.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

tran una $\Delta\chi^2$ significativa ($p < .05$). De nuevo, no se encuentra un cambio notable en CFI ($\Delta CFI < .01$).

Hay que revisar cuándo (en qué ocasión de la prueba) y dónde (en qué ítem y categoría de respuesta) la infracción tiene un impacto sustancial y en qué grado afecta a los cambios en las puntuaciones entre el pretest y el postest. Se realiza un análisis de sensibilidad tal como proponen Liu et al. (2017), cuyos resultados se

presentan en la [Tabla 4](#), que muestra las probabilidades de elección de cada opción de respuesta que predicen los modelos de invarianza métrica e invarianza fuerte.

La violación de la invarianza fuerte produce pequeñas diferencias en las probabilidades de selección de cada una de las opciones de respuesta. Como se puede comprobar, la mayor discrepancia en

Tabla 4
Probabilidades estimadas por los modelos de invarianza métrica e invarianza fuerte

Indicador	Opciones de respuesta (grado de acuerdo)				
	Muy en desacuerdo	2	3	4	Muy de acuerdo
<i>Pretest</i>					
Motivación 2.14	.154 - .154	.149 - .152	.283 - .270	.265 - .282	.149 - .141
Motivación 2.15	.137 - .149	.197 - .187	.277 - .291	.292 - .273	.096 - .100
Motivación 2.16	.111 - .120	.173 - .163	.217 - .262	.332 - .309	.167 - .145
Motivación 2.17	.126 - .130	.172 - .170	.269 - .297	.268 - .254	.165 - .149
Motivación 2.19	.202 - .179	.165 - .191	.330 - .315	.208 - .191	.094 - .123
Motivación 2.21	.141 - .126	.136 - .142	.254 - .210	.291 - .286	.178 - .236
<i>Postest</i>					
Motivación 2.14	.035 - .032	.065 - .059	.182 - .180	.373 - .342	.344 - .387
Motivación 2.15	.031 - .021	.069 - .070	.227 - .224	.415 - .409	.258 - .275
Motivación 2.16	.022 - .017	.059 - .058	.235 - .183	.413 - .410	.271 - .333
Motivación 2.17	.026 - .024	.075 - .069	.263 - .219	.353 - .334	.283 - .353
Motivación 2.19	.020 - .036	.129 - .101	.320 - .323	.294 - .312	.237 - .227
Motivación 2.21	.022 - .027	.084 - .072	.122 - .171	.323 - .345	.449 - .38

Los valores en negrita representan discrepancias superiores a .05 en valor absoluto.

el pretest es de .058 (.236 - .178) y se encuentra en la opción de respuesta 5 del ítem 2.21.

En lo que se refiere al postest, la mayor discrepancia se produce de nuevo en la opción de respuesta 5 de la variable 2.21 (dif. = .063; .449 - .38). Liu et al. (2017) señalan que diferencias inferiores a .05 no deberían afectar sensiblemente a la estimación de los parámetros del modelo de crecimiento de segundo orden. Globalmente, estos resultados indican que el rechazo de la invarianza fuerte no afecta sustancialmente a la elección de una categoría de respuesta específica.

En la Tabla 5 se presentan los índices de ajuste para cada nivel de invarianza factorial longitudinal en la subescala de satisfacción. Los resultados iniciales indican que estos ítems representan el mismo constructo subyacente en cada punto de tiempo. En cuanto al segundo nivel de invarianza (métrica), no hay un aumento significativo de la diferencia $\Delta\chi^2$ ($p = .114$) y un cambio pequeño en el índice de ajuste comparativo CFI ($\Delta\text{CFI} < .01$) respecto al modelo de invarianza configural, que apoya el supuesto de invarianza métrica. El modelo de invarianza fuerte de los datos produce un $\Delta\chi^2$ significativo ($p < .05$) frente al modelo de invarianza métrica, pero solo un incremento leve en el CFI ($\Delta\text{CFI} < .01$). Finalmente, se examina la invarianza estricta. Nuevamente, el modelo de invarianza estricta incrementa la prueba de ajuste χ^2 ; los resultados muestran una $\Delta\chi^2$ significativa ($p < .05$) pero no produce un cambio notable en CFI ($\Delta\text{CFI} < .01$).

Se aplica nuevamente el análisis de sensibilidad propuesto por Liu et al. (2017) con los resultados que se muestran en la Tabla 6. Solo dos de las 120 diferencias son levemente superiores al valor de corte de .05 en el pretest y en postest. En concreto, el porcentaje de estudiantes que el modelo predice que optarán por la opción de respuesta 3 en el ítem 3.26 en el pretest es 5.3% menor (.168 - .221 = .053) en el modelo de invarianza fuerte que en modelo de invarianza métrica. En tanto que, en el postest, el porcentaje de estudiantes que el modelo predice que optarán por la opción de respuesta 5 en el ítem 3.27 es 5.4% mayor (.425 - .371 = .054) en el modelo de invarianza fuerte que en modelo de invarianza métrica. En consecuencia, los resultados generales indican que el rechazo de la invarianza fuerte no afecta sustancialmente a la elección de cada una de las opciones de respuesta.

Modelo de crecimiento de segundo orden (curvas de crecimiento de los factores)

En la Tabla 7 se presentan los índices de ajuste y las estimaciones no estandarizadas de los parámetros del modelo de crecimiento para cada una de las variables latentes. Se obtienen

índices de ajuste adecuados que apoyan un modelo de crecimiento univariante para la motivación de los estudiantes (RMSEA = .045, CFI = .985, TLI = .980, SRMR = .048). En cuanto al modelo relativo al crecimiento de la variable *satisfacción* se obtiene un ajuste razonable RMSEA = .077, CFI = .961, TLI = .962, SRMR = .074). En la Figura 1 se muestran las trayectorias que predice el modelo de crecimiento de segundo orden en cada una de las variables. En estas figuras, la línea gruesa representa la trayectoria promedio, mientras que las líneas delgadas representan trayectorias individuales.

En cuanto a la variable *motivación*, comenzando con los factores de crecimiento, la media del factor de segundo orden relativo a la pendiente (Pend) es significativa ($M = 1.69$, $SE = .134$, $Z = 9.79$, $p < .001$) y representa el cambio medio esperado en el factor de motivación a lo largo del tiempo, de modo que se produce un incremento significativo en esta variable entre el pretest y el postest. El cambio promedio en el postest representa un incremento superior a 1.5 desviaciones estándar frente al pretest. Para determinar la tasa de cambio intraindividual en diferentes puntos en el tiempo, se deben considerar las cargas factoriales de la pendiente (Pend). Las cargas factoriales son 0 en el pretest y 1.315 en el postest ($SE = .19$, $Z = 6.65$, $p < .001$). Se observan, por tanto, diferencias significativas.

Las varianzas de los factores de segundo orden, pendiente (Pend = .60, $SE = .230$, $Z = 2.617$, $p = .009$) e intercepto (Int = 1.09, $SE = .382$, $Z = 2.848$, $p = .004$) son significativas, lo que indica que los estudiantes varían significativamente en su nivel de motivación en el pretest y también en su tasa de crecimiento postest. La covarianza entre los factores de segundo orden, el intercepto y la pendiente, es negativa (-.791). Cuanto mayor motivación inicial (en el pretest) menor tasa de crecimiento.

En lo que se refiere a la variable *satisfacción*, la media del factor de segundo orden relativo a la pendiente (Pend) es significativa ($M = .948$, $SE = .098$, $Z = 9.63$, $p < .001$), de modo que se produce también un incremento significativo en satisfacción entre el pretest y el postest. El cambio promedio en el postest representa un incremento cercano a una desviación estándar frente al pretest. Las cargas factoriales son 0 en el pretest y 1.267 (est) en el postest ($SE = 0.184$, $Z = 6.88$, $p < .001$). Se observan por tanto diferencias significativas.

Las varianzas de los factores de segundo orden, pendiente (Pend = .944, $SE = 0.104$, $Z = 9.050$, $p < .001$) e intercepto (Int = 1.266, $SE = .348$, $Z = 3.635$, $p < .001$) son también significativas, lo que indica que los estudiantes varían significativamente en su nivel de satisfacción en el pretest y también en su tasa de crecimiento postest. La covarianza entre los factores de segundo orden, el intercepto y la pendiente es negativa (-.757). Por lo tanto, al igual que respecto a la motivación, cuanto mayor satisfacción inicial menor tasa de crecimiento.

Tabla 5
Análisis de la invarianza factorial longitudinal de la variable satisfacción

Satisfacción	Invarianza configural	Invarianza métrica	Invarianza fuerte	Invarianza estricta
χ^2 / gl	196.168 / 47	203.934 / 52	257.720 / 69	281.593 / 75
$p \chi^2$	0	0	0	0
$\Delta\chi^2 / \Delta gl$	-	7 / 5	54 / 17*	24 / 6*
$p(\Delta\chi^2)$	-	.114	< .001	.001
CFI	.969	.969	.961	.958
ΔCFI	-	0	.008	.003
TLI	.957	.960	.963	.963
RMSEA (IC 90%)	.082 (.071 - .095)	.079 (.068 - .091)	.077 (.067 - .087)	.077 (.067 - .086)
$p \text{ RMSEA} \leq .05$	0	0	0	0
SRMR	.072	.072	.074	.075

CFI: índice de ajuste comparativo, gl: grados de libertad, IC: intervalo de confianza, RMSEA: error cuadrático medio de aproximación, SRMR: raíz del residuo estandarizado medio, TLI: Tucker–Lewis index, ΔCFI : diferencia entre el índice de ajuste comparativo del modelo y el anterior menos restrictivo, Δgl : diferencia entre los grados de libertad de un modelo y el anterior menos restrictivo, $\Delta\chi^2$: diferencia del estadístico χ^2 entre un modelo y el anterior menos restrictivo.

Un asterisco (*) junto a $\Delta\chi^2$ indica que la diferencia con el ajuste del modelo especificado previamente es estadísticamente significativa.

* $p < .05$

Tabla 6
Probabilidades estimadas por los modelos de invarianza métrica e invarianza fuerte

Indicador	Opciones de respuesta (grado de acuerdo)									
	Muy en desacuerdo		2		3		4		Muy de acuerdo	
	Métr.	Fuer.	Métr.	Fuer.	Métr.	Fuer.	Métr.	Fuer.	Métr.	Fuer.
<i>Pretest</i>										
Satisfacción 3.22	.061	.084	.125	.103	.257	.258	.373	.367	.183	.187
Satisfacción 3.23	.061	.076	.143	.130	.267	.245	.344	.351	.186	.198
Satisfacción 3.24	.114	.127	.148	.128	.288	.262	.284	.282	.166	.200
Satisfacción 3.25	.063	.074	.115	.112	.212	.213	.370	.377	.240	.224
Satisfacción 3.26	.100	.119	.117	.104	.221	.168	.268	.299	.294	.310
Satisfacción 3.27	.150	.152	.113	.116	.192	.218	.226	.246	.318	.269
<i>Postest</i>										
Satisfacción 3.22	.024	.015	.022	.037	.170	.165	.432	.426	.352	.357
Satisfacción 3.23	.013	.009	.035	.043	.157	.171	.459	.447	.336	.331
Satisfacción 3.24	.038	.032	.055	.069	.209	.227	.364	.363	.334	.309
Satisfacción 3.25	.011	.010	.031	.034	.122	.116	.420	.394	.417	.445
Satisfacción 3.26	.031	.024	.035	.043	.063	.106	.343	.308	.529	.519
Satisfacción 3.27	.041	.044	.069	.064	.202	.178	.317	.288	.371	.425

Fuer: modelo de invarianza fuerte, Métr.: modelo de invarianza métrica.
Los valores en negrita representan discrepancias superiores a .05 en valor absoluto.

Tabla 7
Parámetros estimados del modelo de crecimiento de segundo orden

	β		Media	Varianza		Cov.	χ^2 / gl	CFI	TLI	RMSEA (IC 90%)	SRMR
	β_1	β_2	Pend.	Int.	Pend.	Int.-Pend.					
Mot.	0	1.315	1.69	1.087	.601	-.791	156 / 66	.995	.995	.054 (.043-.065)	.045
Satis.	0	1.267	.948	1.266	.944	-.757	257 / 68	.961	.962	.077 (.067-.087)	.074

CFI: índice de ajuste comparativo, Cov: covarianza, gl: grados de libertad, IC: intervalo de confianza, Int: intercepto, Mot: motivación, Pend: pendiente, RMSEA: error cuadrático medio de aproximación, Satis: satisfacción, SRMR: raíz del residuo estandarizado medio, TLI: Tucker-Lewis index.

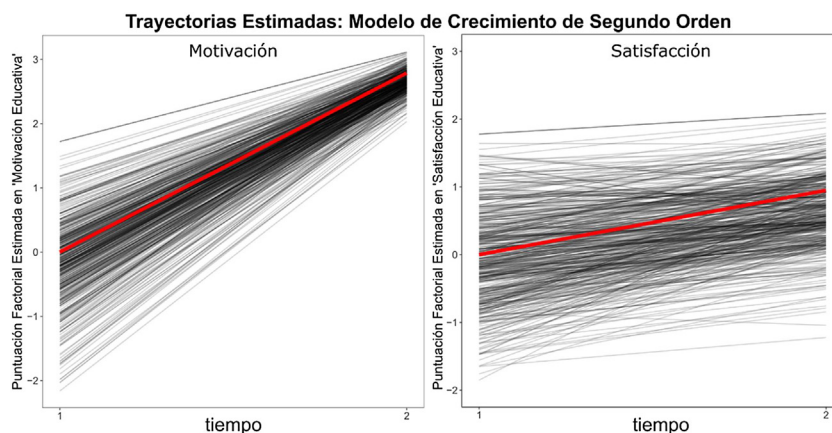


Figura 1. Trayectorias de crecimiento esperadas para las variables motivación y satisfacción.

Discusión

Los resultados muestran una mejora en la valoración de las clases de historia por parte del alumnado de Educación Secundaria tras la implementación de las unidades didácticas basadas epistemológicamente en competencias de pensamiento histórico y en un cambio de metodología didáctica. Tras la aplicación del programa de formación, se observa un aumento sustancial en la motivación de los estudiantes, que es uno de los hándicaps más negativos de la percepción de las clases de historia según muchos estudios (Nokes, 2017; Van Boxtel y van Drie, 2012). Tras la implementación de las unidades, el alumnado expresa que las clases les motivan con el fin de aprender y esforzarse para conocer más sobre la historia. También destaca el hecho de que se incremente la motivación del alumnado porque puede aportar sus propias opiniones y apreciaciones en el aula. Unos resultados similares a otros estudios que se han llevado a cabo en otras materias como las ciencias naturales (Maloy et al., 2019), donde el incremento de motivación es transversal para todos los estudiantes. También se observa un aumento en la satisfacción del alumnado tras la implementación del programa de formación para el profesorado. Este resultado sugiere que el uso en el aula de métodos activos, los cambios en la forma de abordar los objetivos de aprendizaje, el papel del profesorado y el manejo del clima en el aula conducen a una mejora sensible en la satisfacción del alumnado, como comprueba también Reisman (2012) con el uso de fuentes primarias en el aula. Esta autora comprueba una mejora significativa de la comprensión de estas fuentes primarias por el alumnado, asociada también a mayores niveles de satisfacción y motivación, tras una intervención en seis centros de Educación Secundaria de San Francisco.

La percepción mostrada es que el programa formativo aumenta también la satisfacción de los estudiantes respecto a las nuevas formas de trabajo en el aula, ya que todavía siguen dominando estrategias de manejo de la clase de historia tradicionales (Merchán, 2011). El hecho de que la curva de crecimiento de satisfacción tenga menos pendiente que la de motivación se debe a la situación previa. Antes de la intervención el alumnado tiene una motivación media-baja. No obstante, su satisfacción con la asignatura de historia era media-alta. Esta aparente contradicción ya se ha detectado en otros trabajos como los de Fuentes (2004). El alumnado está desmotivado por la metodología que utiliza el profesorado en el aula. Sin embargo, tiene una valoración positiva de la historia como materia formativa y de la utilidad del conocimiento histórico. El cambio metodológico propuesto por las unidades formativas supone una mejora muy sensible en la motivación y satisfacción del alumnado hacia las clases de historia y hacia los procedimientos de enseñanza.

Recientemente Tuithof et al. (2019) han realizado un análisis sistemático de las investigaciones sobre conocimiento didáctico del contenido del profesorado en educación histórica. Los autores destacan el amplio número de estudios cualitativos con muestras pequeñas. También indican que gran parte de los estudios se centran en estrategias disciplinares de la historia como la argumentación o el uso de fuentes primarias en el aula (Ledman, 2015; Monte-Sano y Budano, 2013). No son muchos los trabajos que se adentran en cuestiones de metodología didáctica y, cuando lo hacen, se enfocan desde algún aspecto concreto de la materia, como por ejemplo la pedagogía crítica en el aula (Blevins et al., 2020).

Además, la mayoría de las investigaciones que evalúan programas de formación centran sus resultados en análisis de ejercicios desarrollados por esos docentes, la observación directa, o en cuestionarios de percepción (De Groot-Reuvekamp et al., 2018; Gómez et al., 2019). Pocos son los estudios que han evaluado la mejora de las competencias de ese profesorado en formación inicial cuando acceden a la práctica docente, comprobando la efectividad de las actividades programadas. Y cuando se ha hecho, se realiza en muestras pequeñas y con técnicas cualitativas (Domínguez-

Almansa y López Facal, 2017). Este estudio exploratorio sienta las bases para investigaciones futuras más amplias, que permitan evaluar la adquisición de competencias docentes y los efectos de los programas formativos en educación histórica combinando las técnicas cualitativas con los que habitualmente se realizan, con otras técnicas cuantitativas y de medición sistemática de resultados.

Limitaciones

No ha sido posible aislar el efecto de novedad del programa de intervención por lo que sería deseable comprobar el efecto de la metodología y los recursos empleados cuando estos se prolongan en el tiempo. Debido a las características del diseño, sin grupo de control, no es posible controlar los efectos de historia, maduración y regresión a la media, lo que supone una amenaza a la validez interna. Es por tanto deseable aplicar un diseño que incluya un grupo de control en futuras investigaciones. La ausencia de un grupo de comparación no elimina la posibilidad de una relación causal entre la intervención y la mejora en los resultados, pero el grupo de control es necesario para garantizar las afirmaciones sobre la existencia y la fuerza de dicha relación. El análisis de significatividad práctica de la violación de los diversos niveles de invarianza propuesto por Liu et al. (2017) debe ser investigado más a fondo para confirmar que es igualmente sensible a las violaciones de los diferentes niveles de invarianza (Flens et al., 2019).

Financiación

Este artículo ha sido posible gracias al proyecto de investigación financiado por el Ministerio de Ciencia, Innovación y Universidades, cofinanciado con fondos FEDER, PGC2018-094491-B-C33, y al proyecto «Competencias docentes y métodos activos de aprendizaje. Una investigación evaluativa con el profesorado en formación de ciencias sociales» (20638/JLI/18), financiado por la Fundación Séneca [35500/20638/JLI/18]; Ministerio de Ciencia, Innovación y Universidades [11550/PGC2018-094491-B-C33].

Anexo. Material adicional

Se puede consultar material adicional a este artículo en su versión electrónica disponible en doi: [10.1016/j.psicod.2020.07.002](https://doi.org/10.1016/j.psicod.2020.07.002).

Referencias

- Bangdiwala, S. (1987). Using SAS software graphical procedures for the observer agreement chart. *Proceedings of the SAS Users Group International Conference*, 12, 1083–1088.
- Barnes, N., Fives, H., y Dacey, C. (2017). U. S. teachers' conceptions of the purposes of assessment. *Teaching and Teacher Education*, 65, 107–116. <http://dx.doi.org/10.1016/j.tate.2017.02.017>
- Blevins, B., Magins, K., y Salinas, C. (2020). Critical historical inquiry: The intersection of ideological clarity and pedagogical content knowledge. *The Journal of Social Studies Research*, 44, 35–50. <http://dx.doi.org/10.1016/j.jssr.2019.09.003>
- Browne, M. W., y Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21, 230–258. <http://dx.doi.org/10.1177/0049124192021002005>
- Carretero, M., y van Alphen, F. (2014). Do master narratives change among high school students? A characterization of how national history is represented. *Cognition & Instruction*, 32(3), 290–312. <http://dx.doi.org/10.1080/07370008.2014.919298>
- Chen, F. F. (2007). Sensitivity of goodness of fit indices to lack of measurement invariance. *Structural Equation Modeling*, 14, 464–504. <http://dx.doi.org/10.1080/10705510701301834>
- Cheung, G. W., y Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233–255. http://dx.doi.org/10.1207/S15328007SEM0902_5
- Cuenca, J. M., Estepa, J., y Martín, M. (2017). Patrimonio, educación, identidad y ciudadanía. Profesorado y libros de texto en la educación obligatoria. *Revista de Educación*, 375, 136–159. <http://dx.doi.org/10.4438/1988-592X-RE-2016-375-338>
- Da Rocha, L., Sandro, A., y de Melho, I. (2016). Effectiveness of gamification in the engagement of students. *Journal of Computer in Human Behavior*, 58, 48–63. <http://dx.doi.org/10.1016/j.chb.2015.11.02>

- De Groot-Reuvekamp, M., Ros, A., y van Boxtel, C. (2018). A successful professional development program in history: What matters? *Teaching and Teacher Education*, 75, 290–301. <http://dx.doi.org/10.1016/j.tate.2018.07.005>
- Domínguez-Almansa, A., y López Facal, R. (2017). Patrimonios en conflictos. Competencias cívicas y formación profesional en educación primaria. *Revista de Educación*, 375, 86–109. <http://dx.doi.org/10.4438/1988-592X-RE-2016-375-336>
- Flens, G., Smits, N., Terwee, C. B., Pijck, L., Spinhoven, P., y de Beurs, E. (Publicación online 2019). *Practical significance of longitudinal measurement invariance violations in the Dutch-Flemish PROMIS item banks for depression and anxiety: An illustration with ordered-categorical data Assessment*, <http://dx.doi.org/10.1177/1073191119880967>
- Fuentes, C. (2004). Concepciones de los alumnos sobre la historia. *Enseñanza de las Ciencias Sociales*, 3, 75–83.
- Gómez, C. J., Miralles, P., Rodríguez-Medina, J., y Maquilón, J. (2020). Perceptions on the procedures and techniques for assessing history and defining teaching profiles. Teacher training in Spain and the United Kingdom. *Educational Studies*, <http://dx.doi.org/10.1080/03055698.2019.1707069>
- Gómez, C. J., Monteagudo, J., Moreno, J. R., y Sáinz, M. (2019). Effects of a gamification and flipped-classroom program for teachers in training on motivation and learning perception. *Education Sciences*, 9(4), 299. <http://dx.doi.org/10.3390/educsci9040299>
- Grever, M., Peltzer, B., y Haydn, T. (2011). High school students' views on history. *Journal of Curriculum Studies*, 43(2), 207–229. <http://dx.doi.org/10.1080/00220272.2010.542832>
- Hu, L., y Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. <http://dx.doi.org/10.1080/10705519909540118>
- Ledman, K. (2015). Navigating historical thinking in a vocational setting: Teachers interpreting a history curriculum for students in vocational secondary education. *Journal of Curriculum Studies*, 57, 77–93. <http://dx.doi.org/10.1080/00220272.2014.984766>
- Lesh, B. A. (2011). *Why won't you just tell us the answer? Teaching historical thinking in grades 7–12*. Stenhouse.
- Liu, Y., Millsap, R. E., West, S. G., Tein, J. Y., Tanaka, R., y Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods*, 22(3), 486–506. <http://dx.doi.org/10.1037/met0000075>
- López, C., Carretero, M., y Rodríguez-Moneo, M. (2014). Conquest or reconquest? Students' conceptions of nation embedded in a historical narrative. *Journal of the Learning Sciences*, 24(2), 252–285. <http://dx.doi.org/10.1080/10508406.2014.919863>
- Maloy, J., Fries, L., Laski, L., y Ramírez, G. (2019). Seductive details in the flipped classroom: The impact of interesting but educationally irrelevant information on student learning and motivation. *CBE-Life Sciences Education*, 18(3), 42. <http://dx.doi.org/10.1187/cbe.19-01-0004>
- McArdle, J. J. (1988). Dynamic but structural modeling of repeated measures data. En J. R. Nesselrode y R. B. Cattell (Eds.), *The handbook of multivariate psychology* (2nd ed, pp. 561–614). Springer.
- McDonald, R. P. (2013). *Test theory: A unified treatment*. Psychology Press.
- Merchán, J. (2011). *Práctica de enseñanza y gobierno de la clase*. *Revista Española de Pedagogía*, 250, 521–535.
- Miralles, P., Gómez, C. J., Arias, V., y Fontal, O. (Publicación online 2019). Digital resources and didactic methodology in the initial training of History teachers. *Comunicar*, <http://dx.doi.org/10.3916/C61-2019-04>
- Monte-Sano, C. (2011). Learning to open up history for students: Preservice teachers' emerging pedagogical content knowledge. *Journal of Teacher Education*, 62(3), 260–272. <http://dx.doi.org/10.1177/0022487110397842>
- Monte-Sano, C., y Budano, C. (2013). Developing and enacting pedagogical content knowledge for teaching history: An exploration of two novice teachers' growth over three years. *Journal of the Learning Sciences*, 22, 171–211. <http://dx.doi.org/10.1080/10508406.2012.742016>
- Monte-Sano, C., de la Paz, S., y Felton, M. (2014). *Reading, thinking and writing about history. Teaching argument writing to diverse learners in the common core classroom grades 6–12*. Teacher College Press.
- Muthén, L. K., y Muthén, B. (2015). *Mplus. The comprehensive modelling program for applied researchers: User's guide*. Muthén & Muthén.
- Nokes, J. D. (2017). Historical reading and writing in secondary school classrooms. En M. Carretero, S. Berger, y M. Grever (Eds.), *Palgrave handbook of research in historical culture and education* (pp. 553–572). Palgrave MacMillan.
- Core Team, R. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Recuperado de: <https://www.R-project.org/>.
- Reisman, A. (2012). «Reading like a historian»: A document-based history curriculum intervention in urban high schools. *Cognition and Instruction*, 30(1), 86–112. <http://dx.doi.org/10.1080/07370008.2011.634081>
- Rodríguez-Medina, J., Gómez, C. J., Miralles, P., y Aznar, I. (2020). An evaluation of an intervention programme in teacher training for geography and history: A reliability and validity analysis. *Sustainability*, 3124. <http://dx.doi.org/10.3390/su12083124>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. Recuperado de: <http://www.jstatsoft.org/v48/i02/>.
- Seixas, P., y Morton, T. (2013). *The big six historical thinking concepts*. Nelson.
- Tuithof, H., Logtenberg, A., Bronkhorst, L., van Drie, J., Dorsman, L., y van Tartwijk, J. (2019). What do we know about the Pedagogical Content Knowledge of History teachers: A review of empirical research. *Historical Encounters: A Journal of Historical Consciousness, Historical Cultures, and History Education*, 6(1), 72–95.
- Valls, R., y Colomer, J.C. (2018). El modelo de historias generales. En V. Peris, D. Parra, y X. M. Souto (Coords), *Repensamos la geografía e historia para la educación democrática* (pp. 23–36). Nau llibres.
- Van Boxtel, C., y van Drie, J. (2012). «That's in the time of the romans!» Knowledge and strategies students use to contextualize historical images and documents. *Cognition and Instruction*, 30(2), 113–145. <https://doi.org/10.1080/07370008.2012.661813>
- VanSledright, B. A. (2014). *Assessing historical thinking and understanding Innovative designs for new standards*. Routledge.