Revista de Psicodidáctica 26 (2021) 45–52

Revista de Psicodidáctica

www.elsevier.es/psicod

Original

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

Cosme J. Gómez Carrasco[a], Jairo Rodríguez-Medina[b,*], Pedro Miralles Martínez[a], and Víctor B. Arias González[c]

[a] *Universidad de Murcia*
[b] *Universidad de Valladolid*
[c] *Universidad de Salamanca*

ARTICLE INFO

Article history:
Received 18 May 2020
Accepted 23 July 2020
Available online 8 September 2020

Keywords:
History education
Didactic methodology
Teacher training
Motivation
Satisfaction

ABSTRACT

History classes in Secondary Education form a teaching and learning space with practices that are nourished by two traditional epistemological and methodological models. To counter these models, a training program was carried out in the specialty of Geography and History of the Teacher Training Master's aimed at changing their methodological and epistemological concepts and the design of innovative training units. The effects are evaluated through the implementation of these training units in curricular practices. This evaluation was with a pre-test and post-test formed by two scales (motivation and satisfaction). The analysis strategy was developed in two phases. In the first, the longitudinal factorial invariance is checked, progressively analyzing this invariance between the pre-test and post-test scales. In the second, a second-order growth model or factor curve is applied to assess the change in latent variables between the pre-test and the post-test. After the application of the program, a substantial increase in student motivation and satisfaction was observed. This increase was most visible in motivation, due to the low initial levels of the students.

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

Palabras clave:
Educación histórica
Metodología didáctica
Formación del profesorado
Motivación
Satisfacción

RESUMEN

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.

☆ Please cite this article as: Gómez Carrasco CJ, Rodríguez-Medina J, Martínez PM, Arias González VB. Efectos de un programa de formación del profesorado en la motivación y satisfacción de los estudiantes de historia en enseñanza secundaria. Revista de Psicodidáctica. 2021;26:45–52.

* Corresponding author.
E-mail address: jairo.rodriguez.medina@uva.es (J. Rodríguez-Medina).

Introduction

Traditionally, Secondary Education History classes form a teaching and learning space with generally assumed practices by its members. The most successful students adopt the forms expected by this community to read, memorize, think and write as they are required to (Nokes, 2017). The activities proposed in history classes, the regular interaction with teachers, and the evaluation procedures and techniques proposed by the teacher reinforce the reproduction of these practices (Gómez et al., 2020). These practices are nourished by two models, one epistemological and the other methodological. The first is the General Histories model, an approach to teaching history outlined in the 19th century, from which the aim is to show the totality of historical events chronologically (Valls & Colomer, 2018). The second comes from a methodological tradition that understands teaching practice as a mere reproduction of knowledge built up outside the classroom (Merchán, 2011).

For several decades, research from the area of historical education has sought to reinvent these models, although placing more emphasis on the transformation of epistemological conceptions than on educational practices in the classroom. Thus, Monte-Sano (2011); Reisman (2012); Van Boxtel and Van Drie (2012), are related to issues such as historical thinking, historical literacy or the use of primary sources in the classroom. From another approach, Carretero and Van Alphen (2014); Grever et al. (2011) and López et al. (2014) have focused on issues of historical awareness, identity and historical memory. These studies aim to reinvent the role of the student in History classes: that students understand that learning this discipline is not simply retaining a canonized narrative in memory (Monte-Sano et al., 2014). Learning history involves a different cognitive work that allows for the construction of interpretations of the past based on sources and evidence (VanSledright, 2014). In contrast to what is usually done in the classroom, where texts are used to transmit information to students, who usually memorize them, historians see texts as evidence to build explanations about historical processes (Lesh, 2011).

In order to improve historical education, teachers need to opt for alternative teaching methodologies, but these must be accompanied by an epistemological change. Teachers and students must change the epistemological burdens that conceive history as a set of closed knowledge, which is why it is so important to intervene in teacher training with the aim of improving knowledge, skills and teaching competences related to these methodological and epistemological fields (Barnes et al., 2017).

The present study

In recent years there has been an increase in the analyses of methods and strategies in history classes, with the focus on the use of digital resources (Miralles et al., 2019). In Spain, empirical research in historical education has focused mainly on diagnostic analysis, especially with textbooks, exams, and teacher and student conceptions (Cuenca et al., 2017). Evaluative research, however, is not abundant. The objective of this paper is to analyze the effects of a teacher training program on the motivation and satisfaction of secondary education students (Maloy et al., 2019). The following questions are posed: (1) How does student motivation change with the incorporation of training units?; and, (2) How does student satisfaction change with the incorporation of training units?

Method

Participants

Participants were 473 secondary school students from 18 classes in the Region of Murcia (52% women). The 18 classes are located in 14 schools, 13 of them state and one private. The didactic units were implemented in the four years of compulsory education and the last two years. The choice of participants in the sample is for convenience related to the assignment of the curricular practice centers to the teachers in training who design and implement the training units.

Instruments

A questionnaire comprising two subscales (motivation and satisfaction) was designed following other studies that have evaluated intervention programs based on active methods (Da Rocha et al., 2016). Content validation was by interjudge procedure of the categories of relevance and clarity of the items in the tool, and through a discussion group with seven experts. The Delphi method was chosen and, after the pertinent modifications, a second round was held with the experts to definitively validate the two tools, divided into two subscales. Good agreement results are obtained using the Bangdiwala weighted agreement coefficient (Bangdiwala, 1987): .86 for item clarity, and .91 for item relevance.

The first subscale, related to student motivation, is comprises 6 items grouped into intrinsic motivation (C̈lasses motivate me to know more about history¨) and extrinsic motivation (Ḧistory classes motivate me to get better grades¨). The second subscale refers to student satisfaction and contains 6 items. Example of these items are: Ï am satisfied with the role I have played as a studentör Ï am satisfied with the way the teacher has dealt with the subject.̈ The questionnaire also includes information on the students: age, gender and academic year. A Likert scale of 1–5 was used ranging from (1) v̈ery much in disagreementänd (5) v̈ery much in agreement.̈

The internal consistency of the two subscales and the total score was checked by Cronbach's alpha coefficients for ordinal data and McDonald's omega (McDonald, 2013). Overall, ordinal alpha values of .92 in the pre-test and .92 in the post-test are obtained, and overall reliability coefficients were .91 in the pre-test and .92 in the post-test (Rodríguez-Medina et al., 2020). Both values are considered excellent. On the student motivation scale, adequate reliability indices are obtained (pretest, $\alpha = .89$, $\omega = .90$; posttest, $\alpha = .89$, $\omega = .92$). Similarly, adequate internal consistency indices are obtained for the student satisfaction scale (pretest, $\alpha = .82$, $\omega = .85$; posttest, $\alpha = .82$, $\omega = .85$).

Procedure

An intervention program was designed for the subject M̈ethods and Resources for teaching Geography, History and History of Artïn the speciality of Geography and History of the Master's Degree in Teacher Training (Rodríguez-Medina et al., 2020). The main aim was to improve the skills of future teachers in designing activities and training units. The training program consisted of eight sessions of four hours each. The first session was devoted to answering the question Ẅhy is it necessary to change the didactic model for teaching Geography and Historyänd it is linked to the previous work on historical competences. The second and third sessions exemplify these methods of inquiry. The next two sessions were devoted to work with primary sources, heritage and digital

Table 1
Methodological and epistemological variables incorporated in the didactic units

Methodological variables	Number	Percentage
Digital resources	17	94.44
Colloquiums and debates	17	94.44
Use of portfolios and class work	17	94–44
Cooperative techniques	12	66.67
Inquiry activities	9	50.00
Direct observation, rubrics or observation scales	7	38.89
Epistemological variables	Number	Percentage
Change/permanence activities	16	88.89
Use of historical documents	15	83.33
Cause/consequences activities	13	72.22
Activities of ethical dimension	10	55.56
Historical perspective activities	3	16.67
Activities of historical relevance	3	16.67

resources. The last three sessions addressed the design of the didactic units, applying the theoretical work to the training unit being implemented.

The reliability of the intervention was evaluated via a checklist containing the twelve strategies and techniques that are mainly worked on in the intervention program (Supplementary document 1). Six of these strategies are of a methodological nature. The other six techniques and strategies are of an epistemological nature, and are related to the historical competencies proposed by Seixas and Morton (2013).

Regarding the methodological variables, most of the teaching units use digital resources, colloquiums and debates, portfolios and class work. Cooperative techniques are used in two thirds of the training units, while research activities are used in half of the interventions. Direct observation is used in only four out of ten cases (Table 1). In the case of epistemological variables, most interventions use work with sources and activities of cause and consequence, and of change and permanence. Activities on the ethical dimension are used in just over half of the training units. Activities on historical perspective and historical relevance are less present.

Data were collected for measurement on two occasions: the first week of the curricular practices and at the end of the training unit. The permission of the ethics committee of the University of Murcia was obtained for the procedure of the investigation (supplementary document 2). Informed consent of the participating Secondary Education students was also obtained (supplementary document 3).

Data analysis

Means and standard deviations of items in the pre-test and post-test were calculated and the Wilcoxon sign rank test was applied to check for differences in student perception. To calculate the effect size, the r statistic was used ($r = Z/\sqrt{N}$, $r > .10$ small effect size, $r > .30$ medium effect size and $r > .50$ large effect size). The analysis strategy then followed two phases. In the first, the suitability of the data for factorial analysis were checked with the Barlett sphericity test ($\chi^2 = 7799.53$, $p < .001$) and the Kaiser-Meyer-Olkin sample suitability index (KMO = .91), and then the longitudinal factorial invariance was checked. The factorial invariance between the pre-test and post-test scales was progressively analyzed. The procedure proposed by Liu et al. (2017) was used to check the longitudinal invariance with ordinal data and to evaluate the practical relevance of the invariance violation through sensitivity analysis.

The adjustment of the different structures obtained by confirmatory factor analysis was compared. All models were estimated by weighted least squares ẄLSMV. Goodness-of-fit was checked using the comparative fit index (CFI) and Tucker-Lewis index (TLI) and the root mean square error of approximation (RMSEA). CFI and TLI indexes above .90 are considered to indicate acceptable degrees of fit and above .95 good (Hu & Bentler, 1999). For RMSEA, values equal to or less than .05 are interpreted as good and less than .08 as acceptable (Browne & Cudeck, 1992; Hu & Bentler, 1999). The recommendations of Chen (2007) and Cheung and Rensvold (2002) are followed, according to which increases of less than .010 in IFC and TLI and decreases of less than .015 in RMSEA suggest that there are no relevant changes in the fit of a model with respect to the next more restrictive one. This is done to establish the relevance of the differences in fit between models.

In the second phase, a second-order growth model or factor growth curve is applied (McArdle, 1988) to assess the change in latent variables between the pre-test and the post-test. This model combines a measurement model and a growth model in a single specification, so that it integrates both into a single underlying structure. All analyses are performed using the *lavaan* package (Rosseel, 2012) in the free statistical software R 3.6.3 (R Core Team, 2020) for the second order growth model and Mplus 7.0 (Muthén & Muthén, 2015) for invariance analysis.

Results

Table 2 presents the means in the pre-test and post-test, the differences between means and the results of the Wilcoxon sign range tests for each of the variables observed on the two measurement occasions. The statistics show a significant increase in the assessment of each item in both subscales in the test performed after the intervention.

In the variables *motivation* ($Z = 9.4$, $p < .001$, $r = .443$) and *satisfaction* ($Z = 8.68$, $p < .001$, $r = .408$) the mean effect size is between moderate and high. In the motivation subscale, the students' assessment of the improvement of their motivation to learn and make an effort, and their motivation to learn more about history, stand out. In the satisfaction subscale, the assessment that the students make of the role of the teachers and the work climate in the classroom is noteworthy.

Longitudinal invariance

Table 3 shows the adjustment rates for each level of longitudinal factor invariance in the motivation subscale. The hypothesis of configural invariance is acceptable, since all six items load positively on a single factor on each measurement occasion. The indices show an adequate model fit. No change is observed in the CFI comparative fit index ($\Delta CFI = 0$) with respect to the configural invariance model. The strong data invariance model also produces a significant $\Delta\chi^2$ ($p < .05$) between metric and scalar invariance, but only a slight change in the IFC ($\Delta CFI < .01$). Finally, strict invariance is examined. The strict invariance model increases the fit test χ^2, as expected, and the results show a significant $\Delta\chi^2$ ($p < .05$). Again, no significant change is found in IFC ($\Delta CFI < .01$).

It is necessary to review when (at what time in the test) and where (in which item and response category) infraction has a substantial impact and to what extent it affects changes in scores between the pre-test and post-test. A sensitivity analysis is performed following Liu et al. (2017), and the results are presented in Table 4, showing the probabilities of choice of each response option predicted by the metric invariance and strong invariance models. Violation of the strong invariance produces small differences in the probabilities of selection of each of the response options. As can be seen, the largest discrepancy in the pretest is .058 (.236 - .178) and is found in answer option 5 of item 2.21.

As far as the post-test is concerned, the largest discrepancy again occurs in response option 5 of variable 2.21 (dif. = .063; .449 - .38). Liu et al. (2017) state that differences lower than .05 should not significantly affect the estimation of the parameters of the second

Table 2
Differences in the pretest and postest scores in each scale

Items Motivation	Pretest *M*	Postest *M*	Dif.	*Z*	*p*	Effect size *r*
The way the unit was presented and developed has motivated me to learn more about history.	3.11	3.92	.81	10.12	< .001	.475
My motivation to learn and make an effort in class has increased.	2.98	3.81	.83	10.94	< .001	.516
My motivation has improved because I have a better understanding of the social and cultural reality I am in contact with.	3.29	3.85	.56	7.54	< .001	.358
The unit has increased my motivation to obtain better grades.	3.21	3.78	.57	7.78	< .001	.365
I have been motivated because I was able to contribute my own views and knowledge.	2.82	3.68	.86	10.01	< .001	.471
I have been motivated because we used different resources tan the textbook (Internet, audiovisuals, historical documents, etc.).	3.21	4.12	.91	10.01	<.001	.471
Satisfaction						
I am satisfied with the role I played as a student	3.51	4.05	0.54	8.68	<.001	.408
I am satisfied with the working atmosphere generated in the classroom	3.44	4.06	0.62	9.17	<.001	.431
I am satisfied with the work of my classmates done in small groups	3.25	3.91	0.66	8.98	<.001	.422
I am satisfied with what I have learnt	3.60	4.20	0.6	9.05	<.001	.426
I am satisfied with the way the teacher handled the unit	3.55	4.32	0.77	9.47	<.001	.445
I am satisfied with my grade	3.41	3.93	0.52	6.74	<.001	.316

Table 3
Analysis of the longitudinal factorial invariance of the motivation variable

Motivation	Configural Invariance	Metric Invariance	Strong Invariance	Strict Invariance
χ^2 / *df*	82.860 / 47	91.470 / 52	167.806 / 69	214.664 / 75
p χ^2	.001	0.001	0	0
$\Delta\chi^2/\Delta$ *df*	–	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 (90 % CI)	.040 (.026–.054)	.040 (.026–.054)	.055 (.045–.066)	.063 (.053–.073)
p RMSEA <= .05	.863	.878	.196	.015
SRMR	.045	.046	.046	.048

Note. df = degrees of freedom; $\Delta\chi^2$ = difference of the χ^2 statistic between one model and the previous less restrictive one; Δ *df* = difference between the degrees of freedom of one model and the previous less restrictive one; CFI = comparative fit index; ΔCFI = difference between the comparative fit index of the model and that of the previous less restrictive one; RMSEA = mean quadratic approximation error; SRMR = standardized root mean residual; TLI = Tucker–Lewis index; CI = confidence interval. An asterisk (*) next to $\Delta\chi^2$ indicates that the difference in the fit of the previously specified model is statistically significant.
*p < .05, ** p < .01, *** p < .001.

Table 4
Estimated likelihoods for the metric invariance and strong invariance models

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

Note. Values in bold represent discrepancies greater than .05 in absolute value.

order growth model. Overall, these results indicate that the rejection of strong invariance does not substantially affect the choice of a specific response category.

Table 5 shows the adjustment rates for each level of longitudinal factor invariance in the satisfaction subscale. Initial results indicate that these items represent the same underlying construct at each time point. For the second level of invariance (metric), there is no significant increase in the difference $\Delta\chi^2$ (p = .114) and there is a small change in the IFC comparative adjustment index (ΔCFI < .01) from the configural invariance model supporting the metric invariance assumption. The strong data invariance model produces a significant $\Delta\chi^2$ (p < .05) versus the metric invariance model, but only a slight increase in the IFC (ΔCFI < .01). Finally, strict invariance is examined. Again, the strict invariance model increases the fit test χ^2, the results show a significant $\Delta\chi^2$ (p < .05) but does not produce a noticeable change in IFC (ΔCFI < .01).

The sensitivity analysis proposed by Liu et al. (2017) is again applied with the results shown in Table 6. Only two of the 120 differences are slightly higher than the cut-off value of .05 in the pre-test and in the post-test. Specifically, the percentage of students

Table 5
Analysis of the longitudinal factorial invariance for the variable satisfaction

Satisfaction	Configural invariance	Metric invariance	Strong invariance	Strict invariance
χ^2 / *df*	196.168 / 47	203.934 / 52	257.720 / 69	281.593 / 75
p χ^2	0	0	0	0
$\Delta\chi^2/\Delta$ *df*	–	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 (90% CI)	.082 (.071–.095)	.079 (.068–.091)	.077 (.067–.087)	.077 (.067–.086)
p RMSEA <= .05	0	0	0	0
SRMR	.072	.072	.074	.075

Note. df = degrees of freedom; $\Delta\chi^2$ = difference of the χ^2 statistic between one model and the previous less restrictive one; Δ *df* = difference between the degrees of freedom of one model and the previous less restrictive one; CFI = comparative fit index; ΔCFI = difference between the comparative fit index of the model and that of the previous less restrictive one; RMSEA = mean quadratic approximation error; SRMR = standardized root mean residual; TLI = Tucker–Lewis index; CI = confidence interval. An asterisk (*) next to $\Delta\chi^2$ indicates that the difference in the fit of the previously specified model is statistically significant.
*$p < .05$.

the model predicts will opt for answer option 3 in item 3.26 in the pretest is 5.3% lower (.168 - .221 = .053) in the strong invariance model than in the metric invariance model. Meanwhile, in the post test, the percentage of students that the model predicts will opt for answer choice 5 in item 3.27 is 5.4% higher (.425 - .371 = .054) in the strong invariance model than in the metric invariance model. Consequently, the overall results indicate that rejection of strong invariance does not substantially affect the choice of each of the response options.

Second order growth model (Growth curves of the factors)

Table 7 presents the adjustment rates and non-standardized estimates of the growth model parameters for each of the latent variables. Appropriate adjustment rates are obtained that support a univariate growth model for student motivation (RMSEA = .045, IFC = .985, TLI = .980, SRMR = .048). As for the model related to the growth of the satisfaction variable, a reasonable adjustment is obtained RMSEA = .077, IFC = .961, TLI = .962, SRMR = .074). Figure 1 shows the trajectories predicted by the second-order growth model for each of the variables. In these figures, the thick line represents the average path, while the thin lines represent individual trajectories.

As for the motivation variable, starting with the growth factors, the mean of the second order slope factor (Slp) is significant ($M = 1.69$, SE = .134, $Z = 9.79$, $p < .001$) and represents the expected mean change in the motivation factor over time, so there is a significant increase in this variable between the pre-test and the post-test. The average change in the post test represents an increase of more than 1.5 standard deviations from the pre-test. To determine the intra-individual rate of change at different points in time, the slope factor loads (Slp) must be considered. The factorial loads are 0 in the pre-test and 1.315 in the post-test (SE = .19, $Z = 6.65$, $p < .001$), and significant differences are observed.

The variances of the second order factors, slope (Slp = .60, SE = .230, $Z = 2.617$, $p = 0.009$) and intercept (Int = 1.09, SE = .382, $Z = 2.848$, $p = .004$) are significant, indicating that students vary significantly in their level of motivation in the pretest and also in their post test growth rate. The covariance between the second order factors, intercept and slope, is negative (-.791). The higher the initial motivation (in the pretest), the lower the growth rate.

With regard to the satisfaction variable, the mean of the second order factor related to the slope (Slp) is significant ($M = .948$, SE = .098, $Z = 9.63$, $p < .001$), so there is also a significant increase in satisfaction between the pre-test and the post-test. The average change in the post test represents an increase of about one standard deviation from the pre-test. The factorial loads are 0 in the pretest and 1.267 (est) in the post test (SE = 0.184, $Z = 6.88$, $p < .001$). Significant differences are therefore observed.

The variances of the second order factors, slope (Slp = .944, SE = 0.104, $Z = 9.050$, $p < .001$) and intercept (Int = 1.266, SE = .348, $Z = 3.635$, $p < .001$) are also significant, indicating that students vary significantly in their level of satisfaction in the pretest and also in their post test growth rate. The covariance between the second-order factors, intercept and slope is negative (-.757). Therefore, as with motivation, the higher the initial satisfaction, the lower the growth rate.

Discussion

The results show an improvement in the rating of history classes by Secondary Education students after the implementation of the didactic units based epistemologically on historical thinking skills and a change in teaching methodology. Following the implementation of the training program, there is a substantial increase in student motivation, which is one of the most negative handicaps of the perception of history classes according to many studies (Nokes, 2017; Van Boxtel & Van Drie, 2012). After the implementation of the units, students express that the classes motivate them to learn and strive to know more about history. Noteworthy too is that students' motivation increases because they can contribute their own opinions and assessments in the classroom. These results are similar to other studies with other subjects such as natural sciences (Maloy et al., 2019), where the increase in motivation is transversal for all students.

There is also an increase in student satisfaction following the implementation of the teacher training program. This result suggests that the use of active methods in the classroom, the changes in the way learning objectives are approached, the role of teachers and the management of the climate in the classroom lead to a noticeable improvement in student satisfaction, as is also proven by Reisman (2012) with the use of primary sources in the classroom. The author finds a significant improvement in students' understanding of these primary sources, associated, too, with higher levels of satisfaction and motivation, after interventions in six secondary schools in San Francisco.

The perception shown is that the training program also increases students' satisfaction with the new ways of working in the classroom, as they still master traditional history class management strategies (Merchán, 2011). The fact that the satisfaction growth curve is less steep than that of motivation is due to the previous situation. Before the intervention, the students had a medium-low motivation. However, their satisfaction with the history course was medium-high. This apparent contradiction has already been detected in other works, such as Fuentes (2004). The students are

Table 6
Estimated likelihoods for the metric invariance and strong invariance models

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

Note. Metr. = metric invariance model; Str = strong invariance model. Values in bold represent discrepancies greater than .05 in absolute value.

Table 7
Estimated parameters of the second order growth rate

			Mean	Variance		Cov.	χ^2 / *df*	CFI	TLI	RMSEA (90 % CI)	SRMR
	β_1	β_2	Slp.	Int.	Slp.	Int. – Slp.					
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

Note. Mot = Motivation; Satis = Satisfaction; Cov = Covariance; df = degrees of freedom; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = mean quadratic approximation error; CI = confidence Interval; SRMR = standardized root mean residual; Slp = slope; Int = intercept.

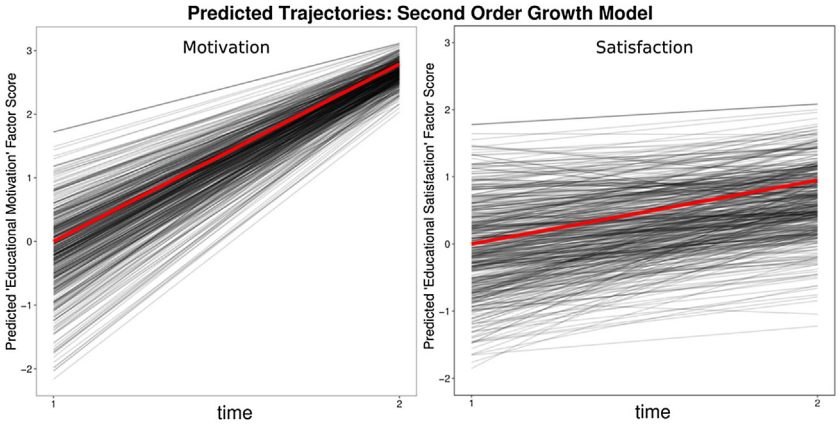

Figure 1. Expected growth trajectories for the variables motivation and satisfaction.

demotivated by the methodology used by the teachers in the classroom. However, they positively rate history as a training subject and the usefulness of historical knowledge. The methodological change proposed by the training units is a clear improvement in the students' motivation and satisfaction regarding history classes and teaching procedures.

Recently, Tuithof et al. (2019) made a systematic analysis of research on didactic knowledge of teacher content in history education. They highlight the large number of qualitative studies with small samples. They also indicate that much of the research focuses on disciplinary strategies of history such as argumentation or the use of primary sources in the classroom (Ledman, 2015; Monte-Sano & Budano, 2013). Not many works go into questions of teaching methodology and, when they do, they focus on some specific aspect of the subject, such as critical pedagogy in the classroom (Blevins et al., 2020).

In addition, most research evaluating training programs focuses on analysis of exercises performed by these teachers, direct observation, or perception questionnaires (De Groot-Reuvekamp et al., 2018; Gómez et al., 2019). Few studies evaluate the improvement of the skills of these teachers in initial training when they enter teaching practice to check the effectiveness of the activities programmed. When this has been done, it is in small samples and with qualitative techniques (Domínguez-Almansa & López Facal, 2017). This exploratory study lays the groundwork for future, broader research, which will make it possible to assess the acquisition of teaching skills and the effects of training programs in historical education by combining qualitative techniques with those that are usually carried out, with other quantitative techniques and systematic measurement of results.

Limitations

It was not possible to isolate the novelty effect of the intervention program, so it would be desirable to check the effect of the methodology and the resources used when these are prolonged. Due to the design characteristics, with no control group, it is not possible to control the effects of history, maturation and regression to the mean, which is a threat to internal validity. Future research should, therefore, apply a design that includes a control group. The absence of a comparison group does not eliminate the possibility of a causal relationship between the intervention and improved outcomes, but the control group is necessary to ensure claims about the existence and strength of such a relationship. The analysis of the practical significance of the violation of different levels of invariance proposed by Liu et al. (2017) should be further investigated to confirm that it is equally sensitive to violations of different levels of invariance (Flens et al., 2019).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at:https://doi.org/10.1016/j.psicoe.2020.08.001.

References

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., & Dacey, C. (2017). U. S. teachers' conceptions of the purposes of assessment. *Teaching and Teacher Education*, *65*, 107–116. https://doi.org/10.1016/j.tate.2017.02.017

Blevins, B., Magins, K., & Salinas, C. (2020). Critical historical inquiry: The intersection of ideological clarity and pedagogical content knowledge. *The Journal of Social Studies Research*, *44*, 35–50. https://doi.org/10.1016/j.jssr.2019.09.003

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*, 230–258. https://doi.org/10.1177/0049124192021002005

Carretero, M., & Van Alphen, F. (2014). Do master narratives change among high school students? A characterization of how national history is represented. *Cognition and Instruction*, *32*(3), 290–312. https://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. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Cuenca, J. M., Estepa, J., & Martín, M. (2017). Patrimonio, educación, identidad y ciudadanía. Profesorado y libros de texto en la educación obligatoria. *Revista de Educación*, *375*, 136–159. https://doi.org/10.4438/1988-592X-RE-2016-375-338

Da Rocha, L., Sandro, A., & De Melho, I. (2016). Effectiveness of gamification in the engagement of students. *Journal of Computer in Human Behavior*, *58*, 48–63. https://doi.org/10.1016/j.chb.2015.11.02

De Groot-Reuvekamp, M., Ros, A., & Van Boxtel, C. (2018). A successful professional development program in history: What matters? *Teaching and Teacher Education*, *75*, 290–301. https://doi.org/10.1016/j.tate.2018.07.005

Domínguez-Almansa, A., & López Facal, R. (2017). Patrimonios en conflictos. Competencias cívicas y formación profesional en educación primaria. *Revista de Educación*, *375*, 86–109. https://doi.org/10.4438/1988-592X-RE-2016-375-336

Flens, G., Smits, N., Terwee, C. B., Pijck, L., Spinhoven, P., & de Beurs, E. (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. Advance online publication.* https://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., & 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.* https://doi.org/10.1080/03055698.2019.1707069

Gómez, C. J., Monteagudo, J., Moreno, J. R., & 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. https://doi.org/10.3390/educsci9040299

Grever, M., Peltzer, B., & Haydn, T. (2011). High school students' views on history. *Journal of Curriculum Studies*, *43*(2), 207–229. https://doi.org/10.1080/00220272.2010.542832

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. https://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. https://doi.org/10.1080/00220272.2014.984766

Lesh, B. A. (2011). *Why wonŕ 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., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods*, *22*(3), 486–506. https://doi.org/10.1037/met0000075

López, C., Carretero, M., & 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. https://doi.org/10.1080/10508406.2014.919863

Maloy, J., Fries, L., Laski, L., & 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. https://doi.org/10.1187/cbe.19-01-0004

McArdle, J. J. (1988). Dynamic but structural modeling of repeated measures data. In J. R. Nesselroade, & 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., & Fontal, O. (2019). *Digital resources and didactic methodology in the initial training of History teachers. Comunicar. Advance online publication.* https://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. https://doi.org/10.1177/0022487110397842

Monte-Sano, C., & 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. https://doi.org/10.1080/10508406.2012.742016

Monte-Sano, C., De la Paz, S., & 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., & 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. In M. Carretero, S. Berger, & M. Grever (Eds.), *Palgrave handbook of research in historical culture and education* (pp. 553–572). Palgrave.

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from 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. https://doi.org/10.1080/07370008.2011.634081

Rodríguez-Medina, J., Gómez, C. J., Miralles, P., & Aznar, I. (2020). An evaluation of an intervention programme in teacher training for geography and history: A reliability and validity analysis. *Sustainability*, *3124* https://doi.org/10.3390/su12083124

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. Retrieved from http://www.jstatsoft.org/v48/i02/.

Seixas, P., & Morton, T. (2013). *The big six historical thinking concepts. Nelson.*

Tuithof, H., Logtenberg, A., Bronkhorst, L., Van Drie, J., Dorsman, L., & 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., & Colomer, J. C. (2018). El modelo de historias generales. In V. Peris, D. Parra, & X. M. Souto (Eds.), *Repensamos la geografía e historia para la educación democrática* (pp. 23–36). Nau llibres.

Van Boxtel, C., & Van Drie, J. (2012). T̈hat's in the time of the romans!Ḱnowledge 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. Innovate designs for new standards. Routledge.*