revista de EDUCACIÓN

Nº 383 JANUARY-MARCH 2019

revista de EDUCACIÓN

Nº 383 January-March 2019
Quarterly Journal
Starting year: 1952

MINISTERIO DE EDUCACIÓN Y FORMACIÓN PROFESIONAL
SECRETARÍA DE ESTADO DE EDUCACIÓN Y FORMACIÓN PROFESIONAL

Instituto Nacional de Evaluación Educativa
Paseo del Prado, 28, 4.ª planta
28014 Madrid
España

Edita

Subdirección General de Atención al Ciudadano, Documentación y Publicaciones

Catálogo de publicaciones del Ministerio: mecd.gob.es
Catálogo general de publicaciones oficiales: publicacionesoficiales.boe.es

Edición: 2018
NIPO línea: 030-15-019-6
NIPO ibd: 030-15-018-0
ISSN línea: 1988-592X 0034-8082
ISSN papel: 0034-8082
Depósito Legal: M.57/1958

Diseño de la portada: Dinarte S.L.
Maqueta: Solana e hijos, Artes Gráficas S.A.U.

Presentation

Revista de Educación is a scientific journal published by the Ministerio de Educación, Cultura y Deporte. Founded in 1940, and since 1952 called *Revista de Educación*, it has been a privileged witness of the development of education in the last decades, and an acknowledged means for the dissemination of education research and innovation, both from a national and international perspectives. It is currently assigned to the Instituto Nacional de Evaluación Educativa within the Dirección General de Evaluación y Cooperación Territorial and it is published by the Subdirección General de Documentación y Publicaciones of the Ministerio de Educación, Cultura y Deporte.

Each year we publish four issues. Starting next issue (No. 361), the magazine will have three sections: Research, Essays and Education Experiences, all of them submitted to referees. In the first issue of the year there is also an index of bibliography, and in the second number a report with statistic information about the journal process of this period and the impact factors, as well as a list of our external advisors.

From 2006 to the second number of 2012 (May-August 358), *Revista de Educación* was published in a double format, paper and electronic. The paper edition included all the articles in the especial section, the abstracts of articles pertaining to the rest of sections, and an index of reviewed and received books. The electronic edition contains all articles and reviews of each issue, and it is available through this web page (www.mecd.gob.es/revista-de-educacion/), where it is possible to find more interesting information about the journal. From the 358 number *Revista de Educación* becomes exclusively an online publication.

Revista de Educación assesses, selects and publishes studies framed in well established lines of research, mainly: methodologies of education investigation and assessment; analysis of education systems and public policies; evolution and history of contemporary education systems; education reforms and innovations; quality and equity in education; curriculum; didactics; school organization and management; attention to diversity and inclusive education; educational guidance and tutorship; teacher selection, training and professional development; international cooperation for the development of education.

Revista de Educación is available through the following data bases:

- *National databases:* ISOC, BEG (GENCAT), PSICODOC, DIALNET, y REDINED (Red de Bases de Datos de Información Educativa).

- *International databases:* Social Sciences Citation Index® (SSCI), Social Scisearch®, scopus, Sociological Abstracts (CSA Illumina), PIO (Periodical Index Online, Reino Unido), IRESIE (México), ICIST (Canadá), HEDBIB (International Association of Universities - UNESCO International Bibliographic Database on Higher Education), SWETSNET (Holanda).

- *Journal evaluation systems*: Journal Citation Reports/Social Sciences Edition (JCR), European Reference Index for the Humanities (ERIH), Latindex (Iberoamericana), scimago Journal & Country Rank (SJR), RESH, Difusión y Calidad Editorial de las Revistas Españolas de Humanidades y Ciencias Sociales y Jurídicas (DICE), CARHUS plus+, Matriu d'Informació per a l'Avaluació de Revistes (miar), Clasificación Integrada de Revistas Científicas (CIRC).

- *Directories*: Ulrich's Periodicals Directory.

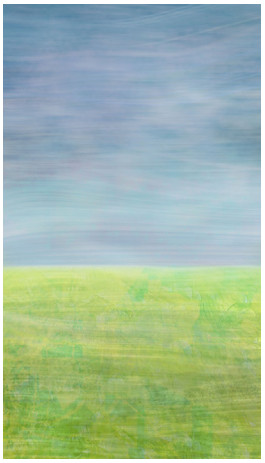
- *National catalogues*: Consejo Superior de Investigaciones Científicas (CSIC-ISOC), Red de Bibliotecas Universitarias (REBIUN), Centro Nacional de Innovación e Investigación Educativa (Ministerio de Educación, Cultura y Deporte), Catálogo Colectivo de Publicaciones Periódicas en Bibliotecas Españolas (Ministerio de Educación, Cultura y Deporte).

- *International catalogues*: WorldCat (USA), Online Computer Library Center (USA), Library of Congress (LC), The British Library Current Serials Received, King's College London, Catalogue Collectif de France (CCFr), Centro de Recursos Documentales e Informáticos de la Organización de Estados Iberoamericanos (OEI), COPAC National, Academic and Specialist Library Catalogue (United Kingdom), SUDOC Catalogue du Système Universitaire de Documentation (France), ZDB Zeitschriftendatenbank (Alemania).

***Revista de Educación* does not necessarily agree with opinions and judgements maintained by authors**

Index

Research



Research

Teacher knowledge and beliefs and their relationship with the purpose of teaching recent history. The transition to democracy in Spain as a controversial case[1]

Conocimientos y creencias del profesorado y su vínculo con las finalidades de la enseñanza de la historia reciente. La transición a la democracia en España como caso controvertido

DOI: 10.4438/1988-592X-RE-2019-383-399

Rosendo Martínez-Rodríguez
Universidad de Valladolid
Carlos Muñoz Labraña
Universidad de Concepción
María Sánchez-Agustí
Universidad de Valladolid

Abstract

This paper considers the link between the knowledge and beliefs of history teachers and their views on the aims and purposes of education, with reference to a controversial period in recent history: the transition from dictatorship to democracy in Spain. For this study, qualitative research was undertaken using interviews with 39 secondary school teachers, grouping their interpretations into three types of thinking: positive-mythical, positive-critical and negative-critical. The results show that most teachers expressed some kind of critical

(1) This study forms part of the TRADDEC research project funded by the Ministry of Science and Innovation in Spain, under their call for I+D+I Research, Development and Innovation projects (Reference: EDU 2009-09775).

Received: 14-03-2017 Accepted: 05-10-2018

interpretation by developing a complex thinking of historical argumentation and analysis. The age variable age did not show remarkable results, but differences were observed with respect to the geographical variable, with the negative-critical view predominating in regions where independence movements were developed. However, in relation to the purposes of history teaching, we find an important distance between the ideal discourse and the concrete examples with which teachers argue their position. That is, many of the teachers with a conception of critical history and connected with the present, found the value of their teaching in an exemplary and closed story of the past, rarely interpretive or critical.

Keywords: history education; teacher beliefs; teacher knowledge; educational aims and values; transition to democracy.

Resumen

En el presente estudio se aborda la vinculación existente entre los conocimientos y creencias del profesorado de historia y sus finalidades y propósitos de enseñanza, sobre un periodo controvertido de la historia reciente: la transición de la dictadura a la democracia en España. Para ello, se ha desarrollado una investigación cualitativa a través de 39 entrevistas a profesores de enseñanza secundaria, agrupando sus interpretaciones en tres clases de pensamiento: *positivo-mítico*, *positivo-crítico* y *negativo-crítico*. Los resultados muestran que la mayoría de los profesores expresaron algún tipo de interpretación crítica desarrollando un pensamiento complejo de argumentación y análisis histórico. La variable edad no arrojó resultados destacables, pero sí se observaron diferencias respecto a la variable territorial, predominando la visión *negativa-crítica* en regiones donde se desarrollan movimientos independentistas. No obstante, en relación con las finalidades de la enseñanza de la historia, encontramos una importante distancia entre el discurso ideal y los ejemplos concretos con los que argumentan su posición. Es decir muchos de los profesores con una concepción de la historia crítica y conectada con el presente, encontraban el valor de su enseñanza en un relato ejemplar y cerrado del pasado, pocas veces interpretativo o crítico.

Palabras clave: enseñanza de la historia; creencias del profesor; conocimientos del profesor; finalidades y valores de enseñanza; transición.

Received: 14-03-2017 Accepted: 05-10-2018

Introducción

Research on teacher thinking has a long tradition in educational research. The interest in knowing about teachers' thinking stems from the notion that their professional behaviour may be influenced by their thought processes, particularly their knowledge and beliefs (Fives & Buehl, 2016). Many previous studies have explored the knowledge and beliefs of teachers, using a variety of methodological approaches (Delaney, 2012; Fives & Buehl, 2016; Grossman, 1990; Kagan, 1992; Shulman, 1987).

In addition to the specific knowledge of teachers, other studies have also focused their attention on the highly subjective feature of thought, on the assumption that there is an ideological web of theories and beliefs that affect teachers' thought and, ultimately, their actual teaching practice (Gimeno & Pérez, 1988). We can define beliefs as a personal form of knowledge, the implicit assumptions that teachers have about students, teaching, classes or content (Kagan, 1992).

Our research deals with the knowledge and beliefs that history teachers have about their subject and how this knowledge and these beliefs relate to the aims and purposes of teaching. As a context for this, we have used a specific historical period: the transition from the Franquist dictatorship to democracy in Spain (starting in 1975 after the dictator's death). This period forms part of the current historiographical, social and political context in Spain and encompasses a number of features that causes this period to remain controversial today.

History is subject to ongoing debate in the field of social science, not only in an academic context, but also in an institutional and social context, and historical discoveries and advances are of value in both public and political spheres, which end up interfering in teaching (Carretero, Berger & Grever, 2017). Studying the teaching of a particular period like the transition to democracy allows us to delve into this framework. A deep historiographical knowledge of this period and the current perspectives on it allow us to establish a coherent classification of teachers' interpretations and beliefs and use this classification to determine the relationship teachers may have with their views on the aims and purposes of education.

The aim of this study, therefore, is to analyse the knowledge and beliefs that history teachers have about a controversial period of recent

Received: 14-03-2017 Accepted: 05-10-2018

history and the relationship this knowledge and these beliefs have with the aims and purposes of teaching. Our research questions were:

- What knowledge and beliefs do history teachers have concerning the transition from dictatorship to democracy in Spain?
- How do they interpret this controversial historical period?
- What aims and purposes do history teachers have when teaching this particular controversial historical period?
- How do the teachers' knowledge, beliefs and interpretations about this historical event impact their conception of the aims and purposes of teaching?

Theoretical background

Conceptions and beliefs about history

Although there have been no previous studies on teachers' knowledge of the transition to democracy, there have been many on the conceptions teachers have of history as a discipline and how these conceptions interact with their teaching.

Evans was a pioneer in this field producing several studies focused on the conceptions that teachers have about history and its meaning (Evans, 1990). He established five categories of conceptions of history, which characterized five kinds of history teachers: storyteller, scientific historian, relativist/reformer, cosmic philosopher and eclectic. Other studies have determined the relationship between the conceptions we have of history, from an epistemological point of view, and the type of historical reasoning teachers expressed (Virta, 2001; Wineburg, 1991). In specific cases, for example, it was found that teachers analysed historical documents in a more simplistic way than would normally be the case in the discipline of history (Bohan & David, 1998, Yeager & Davis, 1996), which could be related to the difficulty their students experience when trying to achieve adequate historical thinking.

Closer to the approach of our research are the studies that have classified teachers' thought according to the type of historical reasoning they profess. For example, Yilmaz (2010), following a long tradition of historiographical debates on the nature of history, proposed two

Received: 14-03-2017 Accepted: 05-10-2018

categories of conceptions of history: "history as the past" and "history as an interpretation of the past". The first represents a closed and "naive" conception of the past, whilst the second is a disciplinary conception. In the study, the author found that many of the teachers were circumventing the interpretative part of history and focused only on the outcome, with obvious repercussions on the way history can be taught.

Similar results were obtained regarding teachers' conceptions of more recent and controversial historical events. In Spain, Sáiz Serrano and López-Facal (2015) explored the kinds of interpretations and narratives produced by students and trainee history teachers about the history of the twentieth century. They concluded that while the teachers demonstrated an increase in critical interpretative narrative compared to the students, exemplary and closed narratives on history still predominated nonetheless.

In Chile (Toledo & Gazmuri, 2010; Toledo, Magendzo, Gutiérrez & Iglesias, 2015), it has been pointed out that, although Chilean professors could have a critical and controversial conception of recent history, as an open story and constructed through diverse interpretations, in their practice they preferred to work little on this type of topics and, when they did it, they mostly maintained neutral positions. Recently, Voet and De Wever (2016) obtained similar results in Belgium, finding that most of the teachers in their study saw history in terms of an interpretation of historical evidence (criterialism) whilst only very few viewed it as a distant and objective description (objectivism) or as a story open to multiple interpretations (subjectivism). The authors concluded that these three distinct approaches also corresponded to different ways of teaching history and that there was also a strong relationship to the specific teaching contexts.

Conceptions and beliefs about the purpose of teaching history

Studies focused on analysing the relationship between conceptions of history and the aims of teaching have been less conclusive but, nevertheless, interesting. According to research obtained by McCrum (2013) using interviews, the way teachers understand the nature of their discipline has an effect on the purpose of their teaching. For McCrum, empiricist conceptions of history are still in the majority and she suggests two types of beliefs about teaching: a teacher-centred approach, based

Received: 14-03-2017 Accepted: 05-10-2018

on the transfer of knowledge, and a student-centred approach, which focuses on the students' own reasoning.

Furthermore, there also appears to be a strong relationship between teachers' conception of history and the aims and objectives of their teaching. On this issue, Husbands, Kitson and Pendry (2003) concluded that teachers with higher and deeper mastery of the content were best able to engage with their students and draw more out of the content. Similar results have also been achieved in very different contexts. For example, in the Philippines, Baraquiel & Oshihara (2014) found that teachers with a broader and more inclusive view of history (called *kasaysayang bayan*) tended to deliver education aimed at the development of historical thinking skills and citizenship education.

Another study, produced in Chile by Toledo and Gazmuri (2010), mentioned above, addressed teachers' purposes in teaching recent history and also identified a greater disparity between teachers and their students. While teachers felt that history should serve their students to understand the present based on the knowledge of the past, students only valued history for the general awareness of culture that they gained from studying the past.

In short, although disciplinary knowledge is not the only or the most important knowledge for teachers, most of the published research indicates its nature and depth are crucial when teaching history.

The transition to democracy in Spain as a controversial case study

It is normal for historical events that are recent and that represent a conflict between parts of a society to produce uncertainty in the way that they should be treated in the field of education. In the most severe cases, this kind of history tends not to be included in the national curriculum or else is delivered using a mythical or imposed narrative, without any opportunity to offer multiple perspectives (Paulson, 2015;).

Much has been published about teaching issues that remain socially current and controversial, and about their didactic treatment and the value of properly introducing discussions on the topic in the classroom (Barton & MacCully, 2007; Legardez & Simonneaux, 2006; Low-Beer, 1999; Psaltis, Carretero & Čehajić-Clancy, 2017). In some cases, teachers have been studied from the point of view of their teaching strategies

Received: 14-03-2017 Accepted: 05-10-2018

and their ideas on controversial issues (Katrin, 2016; Toledo, Magendzo, Gutiérrez & Iglesias, 2015). However, little research has been undertaken on how teachers process this type of historical content or how their knowledge and beliefs can interact with the purposes of their teaching. We share the belief that, when a society is divided over interpretations of the past, the focus and purpose of teaching may become more important than the content that is covered (Cole & Barsalaou, 2006).

In the case of Spain, the most controversial part of its recent history relates to the civil war (1936-1939) and Franco's subsequent dictatorship (1939-1975). The presentation and treatment of these events in curricula and textbooks have evolved in parallel to the social and democratic changes that have occurred. During this development, changes have also been seen in the purpose of teaching history, moving from a system based on promoting nationalist sentiments (López-Facal, 2003) to the current approach of preparing students to interpret the past and analyse society (Prats & Santacana, 2014). In this transformation, the use of active methodologies and the competency-based approach, beyond substantive content, have played a fundamental role and are currently the subject of multiple studies (López-Facal, Miralles, Prats & Gómez, 2017).

We have chosen the case of the transition to democracy in Spain as a controversial case because of its *connection to today* and its controversial political issues (monarchical system, territorial organization, reform of the Constitution, electoral law, etc.); its permanence in the collective memory and the questioning of a policy based on "the pact of silence" or "forgetting" (Espinosa, 2007; Kovras, 2014). Finally, this historical period was chosen because of the historiographical confrontation between the predominant viewpoint that praises the rapid and peaceful nature of the transition (Fusi, 2009; Linz, 1996) and those who accuse it of lacking self-criticism and of maintaining a politically correct and official narrative (Gallego, 2008), Muniesa, 2005).

Methodological design

The qualitative method was chosen as the most suitable strategy to analyse the network of teachers' conceptions and beliefs in depth, instead of the numerous but more superficial results that the quantitative method could offer us. With this purpose, it was decided to carry out a series of

Received: 14-03-2017 Accepted: 05-10-2018

semi-structured interviews, promoting the open development of each question (Kvale, 2011).

The sample

The sample comprised a total of 39 secondary school history teachers. The number was decided with the objective of obtaining wide and varied information, but, at the same time, an in-depth study of each one of the participants was possible. The participants were recruited in cities of seven Spanish autonomous communities, with the aim of representing the centre of the country (Madrid and Valladolid) and, especially, the periphery (Barcelona, Bilbao, Murcia and Santiago de Compostela). This selection is motivated, on the one hand, by the interest in achieving the representation of diverse national geographic and cultural areas, and on the other hand, in including regions in which, due to their historical characteristics or their socio-political evolution, our theme of study could offer different views and experiences, as in the cases of Catalonia and the Basque Country. Moreover, the sample design considered the following characteristics of the participants: a) all teachers were teaching the historical period of the transition in the 4th year of compulsory secondary education or the 2nd year of high school; b) 50% of participants were teachers aged 45 years or older with personal memories of the transition and the rest were teachers younger than 45 years of age; c) a balance was sought between teachers from public schools (60%) and teachers in private or semi-private schools (40%). In regards the academic level of the teachers, all had a bachelor's degree (usually in history, but some in other social sciences) and in some cases they also had specialist qualifications such as master's degrees (10) or doctorates (4).

The interviews

The discursive nature of our research, with its objectives relating to knowledge and beliefs, determined our choice of interviews as the most appropriate instrument in our quest for spontaneous and free responses from the subjects. We selected the semi-structured type of interview from the wide range of interview types available (Flick, 2007). Whilst

Received: 14-03-2017 Accepted: 05-10-2018

maintaining the open character of a qualitative interview, semi-structured interviews also allow items to be pre-designed in the form of a script. This interview method has been commonly used by other research groups to study teachers' thoughts (e.g. Baraquiel & Oshihara, 2014; McCrum, 2013; Voet & Weber, 2016; Yilmaz, 2010)

An initial version of the interview was trialled with teachers from Valladolid and then reviewed by a group of experts. The final interview format agreed on was then applied to all the teachers in our sample, in conversations lasting between 50 and 70 minutes. The interviews were conducted in intimate and quiet environments, respecting the teachers' anonymity in order to generate greater confidence.

The categories that structured the script of the interview were the teachers' knowledge and beliefs about: (a) the content, knowledge, beliefs and interpretation and (b) the aims and purposes of teaching the content.

Analysis of the information

The interviews were recorded, transcribed and then uploaded to the Atlas.ti programme to facilitate their analysis. The transcription of all the interviews produced a total of 242,572 words. In accordance with Kvale (2011), the analysis of meanings was conducted in three phases: encoding of meaning, condensation of meaning and interpretation of meaning. Encoding combined those ideas, concepts, interpretations and propositions, which we used to pursue our research objectives. An initial theoretical coding (top down) was performed and during the course of the analysis new categories emerged (bottom-up). The initial categories (see Table 1) came from reading a wide range of works about the period of the transition to democracy and the thoughts of history teachers.

TABLE I. Initial Encoding

Family	Code	Sub-code
Knowledge, beliefs and interpretations regarding content	Knowledge and beliefs Interpretative historiographical elements	Key historic factors Key events Connection to the present Chronological framework of the period Violence during the transition Pacts and agreements
Aims and purposes of teaching	Type of history and historical thought Values to be developed	

While the initial review of the interviews and the primary analysis was undertaken by the same researcher, a second researcher was responsible for completing a parallel review of the transcripts and further analysis, in order to ensure the validity of the results through interpretive agreement.

Results

To facilitate the presentation of the results, the information is divided according to the two main groupings: knowledge, beliefs and interpretations of content and knowledge, and beliefs about the aims and purposes of education.

Knowledge, beliefs and interpretations about the transition to democracy

To analyse and classify the teachers' knowledge and beliefs, the way they presented the topic was analysed in relation to three key historiographical features of the transition: the violence during the transition, the role of pacts and agreements, and the chronology of the period. In the first case, for one part of the historiography, the violence was not studied to any significant degree, the period of the transition being therefore depicted as exemplary and peaceful (Fusi, 2009; Linz, 1996; Tusell, 1994). For other history teachers, however, violence played a very important role in

Received: 14-03-2017 Accepted: 05-10-2018

the process (Aguilar & Sánchez-Cuenca, 2009; Sánchez, 2010). Something similar happens with the pacts and agreements, which were exemplary for some (Fusi, 2009; Linz, 1996), whilst for others they represented a way of allowing the continuity of the Franco regime and enabling those involved to escape prosecution for their crimes (Muniesa 2005). Finally, the chronology is divided between those who consider that the period ended in 1978 or 1982 (most historians, for example Maravall & Santamaría, 1994; Soto, 2009) and those who believe that even today the process of real transition continues or even that a second transition is needed (Muniesa, 2005).

In addition to historical analysis, teachers were also asked to produce an evaluative interpretation of the period being studied. They were asked if they saw a connection with contemporary political, social or economic affairs. From a combined analysis of their responses and their general interpretations, the results were classified into three basic groups of thought: *positive-mythical, positive-critical* and *negative-critical*. This categorization should not be understood as isolated compartments, but, as in historiography, the teachers' reflections could mix ideas from the three models. However, in the whole construction of their discourse, the teachers tended to place themselves quite clearly in one category or another.

A *positive-mythical* view of history

This group includes all those teachers who view the period as an essentially exemplary stage in the history of Spain, with few critical reflections on the subject. Often they based their view on the notion of the transition as an established model exportable to other realities through which a few political architects have managed the process of change masterfully, peacefully and with the general consensus of the population. Adolfo Suárez and King Juan Carlos appear as protagonists in delivering the pacts that produced change, sometimes with mythical connotations. The chronology of the period ends in 1978 when the Constitution was approved. Of the 39 teachers interviewed, 14 shared that view of history.

One particular case is Teacher 3 who sees the period as a historical product exportable to other realities, a model transition of which they feel proud: "…a situation handled well and that should be exported. It

is natural for any country where a new democracy is created for them to copy Spain. Spain is taken as a model in terms of the role of the King, or the role of political parties" (T.3/-45 years). Note the value-laden words of Teacher 3 and how their conception of history implies a completed and closed product, with no opportunity for critical interpretation.

Another important element in this vision is the personification of historical processes. The political figures involved acquire epic connotations and are seen as singular craftsmen, ignoring other factors of historical analysis such as context or causality. For example, Teacher 15 states: "To me it seems the transition was carried out very well and especially how the King handled things, as he was one of the first who tried to impose democracy, that is to say, impose, in quotation marks" (T.15/+45 years).

These interpretations are charged with beliefs but, nevertheless, are consistent with the model presented by some political studies of the 1990s, which portrayed the case of the Spanish transition as a model (Linz, 1996). They also coincide with the vision presented by large sections of the media, in documentaries and books, during most of the ensuing democracy. For example, the volumes published by El Mundo or the television series of Victoria Prego presented a vision of the period characterized by the role of political elites and with little prominence of other historical elements, such as citizenship (Quirosa-Cheyrouze, 2009).

A *positive-critical* view of history

In this category we grouped together those positions that also interpret the process of the transition in a positive way, but through an open and critical conception of events. In these cases the teachers use examples in their arguments that justify their position through more complex mechanisms of analysis than those employed by the former group. For example, they do not emphasize the peaceful nature of the transition but rather argue that it would have been more violent without the pacts and agreements achieved. The chronology of the period either ends in 1978 or after the failed coup attempt in 1981. This is the largest set of perceptions, representing 17 teachers.

This conception of history incorporates an analysis of the past which acknowledges the context and different historical conditions. One issue

Received: 14-03-2017 Accepted: 05-10-2018

of historical conditioning presented by several interviewed teachers is terrorism, as raised by Teacher 27: "…with the presence of state terrorism and ETA and other terrorist groups, the transition laid very solid foundations for democracy" (T.27/-45 years). Another recurring aspect of historical conditioning is the memory of the civil war, as seen in the analysis of Teacher 13 who says: "It ended a dictatorship that had resulted from a war, in a period in which the protagonists of the war were still alive, and seven years later we had a socialist government" (T.1/+45 years). Both factors, the impact of terrorist violence and the memory of the civil war, are widely repeated by those expressing more moderate historiographical viewpoints (Maravall & Santamaría, 1994).

A *negative-critical* view of history

A third option is that expressed by teachers who are openly critical of both the historical process and its result. They articulate an open and interpretive historical reasoning, as in the previous case, but their arguments and reflections follow very different paths. They tend to focus their discourse on contentious issues such as the non-prosecution of crimes under Franco or the continuation of the regime's politics and economy. They also deliver a critical interpretation based on the contextualization of the events, but also taking into account other interpretive elements. The model of a negotiated transition is now perceived in a negative way, reflected in their historiography (Espinosa, 2007). In addition, coinciding with the most critical historiography, violence becomes more important in their discourse and, according to some, the transition period has not yet ended. A total of 8 of the teachers interviewed share this approach.

In terms of the pacts, Teacher 34, who was born during the Franco dictatorship, argues that, "Franco's regime legitimized democracy, and democracy legitimized the Franco regime" (T.34/+45). For these teachers, the pacts represented an end to the demands of the political left, who gave up their aspirations in order to participate in the process. In the case of Teacher 5, in a reflection highlighting their personal disenchantment, they state "…the mobilization of the people and the organizations that really made a struggle against Franco, what use were they? None, almost nothing"(T.5/-45). Another consequence of the pacts was the continuation of elements and characteristics of the dictatorship into the

democracy. For example, Teacher 33 contends: "There was continuity of ministers, people who were on the front line in the Civil Guard ... who continued in democracy without any explanation" (T.33/-45 years). This argument has been corroborated by historiographical and journalistic research (Grimaldos, 2004) and correlates to the lack of a process for justice during the transition (Kovras, 2014).

It is important to note that the vast majority of the *negative-critical* views were found in teachers from the historic communities of Galicia, the Basque Country and Catalonia, where political and social movements for regionalism or independence are prominent. In many cases, this position is linked only to speeches that are very critical of the process of transition at the national level, neglecting the internal reflection on the distribution of power among pro-independence parties in these communities, also derived from the same transition process. The ideological and geographic factors are therefore presented as variables to take into account when considering teacher thinking about controversial periods in recent history.

Aims and purposes of education and their relation to knowledge, beliefs and interpretations of history

In the interviews, teachers were asked what they felt were the purposes and goals of teaching history and, particularly, in teaching the period of the transition. Usually, these history teachers identify a strong value in the teaching of history as a means for students to better understand the present and to act to change it. However, when we investigated this aspect further, we found a significant discrepancy between this ideal discourse and concrete examples of how they actually support their position. This suggests that many teachers with a critical and connected conception of history still found the value of their teaching in an exemplary and closed version of the past, only rarely becoming interpretive or critical.

Received: 14-03-2017 Accepted: 05-10-2018

TABLE II. Conceptions of history and the purposes of teaching

Conception of history	Knowledge, beliefs and interpretations	History to teach/purpose/ values	N° of teachers
Positive-mythical	Wide coverage of pacts and historical characters. Model transition. Peaceful transition. End of process: 1978.	*History as example.* Understanding the present. Values: consensus, national and constitutional unity, peace, democracy.	14
Positive-critical	Critical interpretation. Historical factors: terrorist violence, memory of the civil war. End of process: 1978/1981.	*History as example/history as problem* Understanding the present. Values: consensus, peace, participation, democracy.	17
Negative-critical	Critical interpretation. Historical factors: lack of justice during the transition, concessions of the left. End of process: 1981/not yet ended.	*History as example/history as problem* Understanding the present. Values: concessions, transitional justice, democracy.	8

As can be seen in Table 2, neither the type of interpretations given by the teachers nor their beliefs and values associated with this historical period are reflected in their views on the aims of education. We can divide the results into two groups: in relation to the type of historical thinking they feel should be developed (*history as example* or *history as problem*) and in relation to the values they associate with their teaching.

History as example or history as problem

When we analyse the interviewed teachers' responses we find two ways to understand the purpose of teaching history, which correspond to two ways of thinking about history: the first and most common, which we have called *history as example*, sees history as a narrative that contains teaching within itself; the second, *history as problem*, has its educational value not in the telling of the story but rather in the analysis or interpretation of past events.

As shown in Table 2, all teachers with a *positive-mythical* conception of history produced responses about the aims of education that are readily classifiable as *history as problem*. For example, if we consider Teacher 15,

Received: 14-03-2017 Accepted: 05-10-2018

discussed above in the section on *positive-mythical* conceptions, we see that their lack of complex historical reasoning coincides with a simplistic and exemplary conception of what they aim to teach: "…if [the students] study the beginnings of democracy and the rules and the laws presented in the Constitution, it is like giving them guidelines to behave in the right way and be good citizens" (T.15/+45).

However, it is surprising to see that a large proportion of the teachers with *positive-critical* and *negative-critical* conceptions also articulated the purpose of teaching in terms of *history as example*. In these cases there is inconsistency between their understanding of history, in a critical and interpretive sense, and the purpose of teaching, as an exemplary account. This inconsistency is justified in some cases because the teachers consider that the history of the transition, being both recent and complex, cannot be studied in depth with students but rather only expressed in very concrete terms. Teacher 21 says, "We must explain the difference between a dictatorship and a democracy, democratic values. But you cannot cover it in depth because you could confuse students, and we must be very careful" (T.21/+45). In other words, the explanation of a closed and simplified story prevents teachers from discussing the recent past, ignoring a good part of its formative value.

Teachers who expressed a more critical view of the transitional past (whether *positive-critical* or *negative-critical*) showed greater complexity in articulating their learning goals, which we define as *history as problem*. In these cases, their aims for teaching coincide with their interpretive and critical conception of history. For example, Teacher 34 states that it is important to teach "…to adopt a critical point of view, to understand the positive and negative aspects of the pacts, agreements, concessions..." (T.34/-45). In addition, the critical connection these teachers make with the present through interpreting history are also apparent in their teaching goals. Students "…must understand what political and legal regime they now live under; and not just the Constitution itself but how the Constitution was created, how it was drafted and debated", according to Teacher 5 (T.5/+ 45).

Received: 14-03-2017 Accepted: 05-10-2018

Values associated with the learning of history

Another important element of our analysis relates to the values that teachers of the study express as part of the aims of their teaching. As shown in Table 2, teachers with a *positive-mythical* conception of history consequently articulated values such as the consensus of the political class, national unity, the peaceful resolution of conflict and the triumph of democracy. For example, Teacher 2 says that the study of the transition serves to teach "…consensus, a new organization of the State, the order of the Constitution and the value of the political class committed to change and to improving the country" (T.2/-45).

In the case of teachers with a *positive-critical* conception of history, they share some values about teaching with the previous group, such as the consensus, the peaceful resolution of conflict and the triumph of democracy. But they also introduce a new value: participation, giving importance to the role of ordinary people and not just the political class. Teacher 27, for instance, says: "nowadays we act more like sheep. Not being exemplary, I think [the transition] was a time of participation, of involvement, of struggle, of being critical as citizens "(T.27/+45).

Positive values are mentioned even by teachers with a *negative-critical* view of history. Understandably, they do not share the values of national unity or consensus, but the value of democracy over dictatorship is highlighted and their critical discourse is limited when talking about the purposes of teaching. Some teachers expressed a *critical-negative* value when teaching the period: the value of justice, which in relation to the transition is seen as a value that failed. Teacher 7 says of this, "… what we learn is that if there is a transition it is best done peacefully, but always with justice. What cannot be right is for the oppressors themselves to remain in power. It seems to me unfair" (T.7/+45).

In short, the teachers' values, beliefs and interpretations follow a line of argument that is transferred from their conceptions of history to the purposes of their teaching.

Discussion

Our classification of teachers' thinking into three categories does exhibit broad similarities with the distinction Yilmaz (2010) makes between

Received: 14-03-2017 Accepted: 05-10-2018

"history as the past" and "history as an interpretation of the past". In our case, however, most of the teachers interviewed talked about history in terms of interpreting the past (17 *positive-critical* conceptions and 8 *negative-critical* conceptions), although there is no small number of teachers with a closed conception of history, as a simple story of the past (14 *positive-mythical* conceptions). These results coincide with the work of Voet and De Wever (2016) if we acknowledge that our *positive-critical* and *negative-critical* classifications can be identified with the criterialist position mentioned in their work.

In terms of the quantitative study conducted in Spain by Sáiz Serrano and López-Facal (2015), there are similarities with our results because many of the teachers present accounts of history that are quite rich in content, non-linear and associated with areas of analysis (social, political, economic, etc.). But, as in the study by Sáiz Serrano and López-Facal, many of the teachers we interviewed did not develop a critical interpretive vision of controversial aspects of the recent past.

Studies in Chile, within a context of recent historical events similar to those in Spain, reveal history teachers that are more critical than those in our research (Toledo & Gazmuri, 2010). However, our results do concur with those of Toledo and Gazmuri in that the teachers overwhelmingly agreed that the purpose of teaching history was that their students should better understand the present. This model of teachers coincides with studies undertaken by Evans (1990) which, even then, highlighted the character of the relativist/reformer teacher, concerned with bringing about social change through their work.

We also agree with the research published by Toledo, Magendzo, Gutiérrez and Iglesias (2015) in which teachers have high expectations when defining their goals for teaching recent history, which then do not correspond to the way in which they propose to achieve them. Many teachers of the study therefore, even those with interpretive and critical conceptions of history, viewed learning history in terms of an exemplary tale of the past, not as a means of interpretation or problematisation that could serve to understand the present. As a result, there is a disparity between the ideal aspiration of these teachers and the means by which they aim to achieve it.

With regard to the values associated with the teaching of history, they represented an important element of the educational aims of the history teachers interviewed. This could be due to the specific characteristics

Received: 14-03-2017 Accepted: 05-10-2018

of the historical period since the transition represents the advent of democracy after a long dictatorship. So, although many teachers gave a critical, even negative, interpretation of the historical period, when articulating their purposes almost all tried to identify some positive values to deliver in their teaching. The role of the teacher as educator (focusing on behaviour patterns and values) is therefore predominant, compared to the role of the expert or specialist (dedicated to teaching discipline knowledge), in line with the classifications identified by Monereo and Monte (2011).

We agree with McCrum (2013) and with Baraquiel and Oshihara (2014) in that there is a relationship between the conceptions of history and the purposes of teaching, but with certain nuances. Unlike the research of Baraquiel and Oshihara (2014), which found that teachers with a broader and more critical understanding of the past had a richer and more complex vision of their teaching (seeking to develop the skills of historical thinking), in our study this was not always the case. Most of the surveyed teachers tended to limit their critical and interpretive conception of history when defining their teaching goals, reducing the analytical burden on the past and basing it more on a closed account.

Conclusions

On knowledge, beliefs and interpretations of history. The teachers of this study displayed a very general knowledge of the recent past which they reinforced with beliefs and values taken from within contemporary society and media. Their knowledge, beliefs and interpretations about the history of the transition corresponded to the historiographical debates presented at the beginning of this article. Based on these factors and on the historical reasoning they professed, the participating teachers were put into three groups: *positive-mythical*, *positive-critical* and *negative-critical*.

Overall, a majority of teachers expressed some kind of critical interpretation (17 *positive-critical* and 8 *negative-critical*), developing a complex line of argument and historical analysis. However, it is equally true that 14 teachers (*positive-mythical* conceptions) only delivered a closed historical account, without developing any real interpretive argument.

Received: 14-03-2017 Accepted: 05-10-2018

The age variable did not return significant results, with teachers older and younger than 45 appearing in each of the categories. However, different results were observed regarding the geographic variable. *Negative-critical* views predominated in regions where regionalist or independence movements are strong and where they tend to be highly critical of the way the transition to democracy come about; while positive interpretations were more apparent in those regions where this problem does not exist. This aspect highlights the impact that current thinking and the media have on the kind of historical interpretations and beliefs these teachers hold.

On the relationship between knowledge, beliefs and interpretations and the aims and purposes of education. In the case of teachers with a *positive-mythical* vision, there is a clear relationship between their vision and the purpose of education. The same closed accounts and the same values intrinsic in their historical interpretations are repeated in their discourses about teaching goals. Their principal concern is the transmission of exemplary history where the story itself offers learning, combined with a number of positive values such as consensus, peace, national unity and democracy.

The case of teachers with critical views of the past, either positive or negative, proved more complex. Although there was a relationship between critical views and deeper and more complex purposes of teaching history, in many cases, when expressing these educational purposes, they omitted any critical discourse or opportunities for more complex interpretation. In short, the history taught seems to simplify and to become an exemplary story, which assumes homogeneous and sometimes superficial values. So a main purpose of teaching history is limited, such as the development of historical thinking and critical thinking skills through multicausal analysis and from multiple perspectives of the past (Prats and Santacana, 2014; Rüsen, 2005; and López-Facal, 2015). Only a few teachers actually placed the focus of their educational purpose on the interpretation and questioning of the past rather than on the historical account itself. These cases were more frequent in relation to the *negative-critical* conceptions, which leads us to believe that their perceptions of the past do influence the purposes of their teaching.

To take this research forward, we believe it is important to extend it by carrying out in-class studies. We need to analyse what actually happens in classrooms where this controversial and recent issue is being

Received: 14-03-2017 Accepted: 05-10-2018

taught; how teachers approach these subjects and if they are consistent with their conceptions, beliefs and purposes. At the same time, our research reveals how teachers face real difficulties in converting their own historical knowledge into the desired learning outcomes for their students. Future research, therefore, should also focus on strategies and materials that enable teachers to further develop the skills of historical thinking.

References

Aguilar, P. & Sánchez-Cuenca, I. (2009). Violencia política y movilización social. In Baby, S., Compagnon, O., & González, E. *Violencia y transiciones políticas a finales del siglo XX* (pp. 95-111). Madrid: Casa de Velázquez.

Baraquiel A. C. & Oshihara, S. (2014). Teachers' Pedagogy and Conceptions of History: Decolonizing and Transforming History in Elementary. *The SIJ Transactions on Industrial, Financial & Business Management (IFBM), 2*(3), 87-94.

Barton, K. C. & McCully, A. W. (2007). Teaching controversial issues... where controversial issues really matter' in Teaching History. *Teaching History. The Historical Association,* 127, 13.

Bohan C. & Davis O. (1998). How social studies student teachers' historical thinking is reflected in their writing of history. *Theory and research in social education,* 26, 173-197.

Buehl, M. M., y Beck, J. (2015). The relationship between teachers' beliefs and practices. En H. Fives y M. Gregorie Gill (Eds.), *International handbook of research on teachers' beliefs* (pp. 66-84). New York, NY: Routledge

Carretero, M., Berger, S. y Grever, M. (2017). Introduction: Historical Cultures and Education in Transition. En Carretero, Berger y Grever (Eds.). *Palgrave Handbook of Research in Historical Culture and Education* (pp. 1-37). Palgrave Macmillan.

Cole, E., & Barsalou, J. (2006). *Unite or Divide? The Challenges of Teaching History in Societies Emerging from Violent Conflict. Special report*. Washington, DC: U.S. Institute of Peace.

Espinosa, F. (2007). De saturaciones y olvidos. Reflexiones en torno a un pasado que no puede pasar. *Hispania Nova. Revista de Historia Contemporánea Hispania Nova*, 7. Retrieved from http://hispanianova.rediris.es/7/dossier/07d013.pdf

Evans, R. W. (1990). Teachers' conceptions of history revisited: Ideology, curriculum and student belief. *Theory and Research in Social Education, 18*(2), 101-138.

Fives, H., & Buehl, M. M. (2016). Teachers' Beliefs, in the Context of Policy Reform. *Policy Insights from the Behavioral and Brain Sciences,* 3(1), 114–121.

Flick, U. (2007). *Introducción a la investigación cualitativa*. Madrid: Morata.

Fusi, J. P. (2009). España 1975-2008. In Carr, R. *España 1808-2008* (epilogue). Barcelona: Ariel.

Gallego, F. (2008). *El mito de la transición. La crisis del franquismo y los orígenes de la democracia (1973-1977).* Barcelona: Crítica.

Gimeno, J., & Pérez, A. (1988) Pensamiento y acción en el profesor: de los estudios sobre la planificación al pensamiento práctico. En *Infancia y Aprendizaje*, 42, 37-63.

Grimaldos, A. (2004). *La sombra de Franco en la transición*. Madrid: Oberon.

Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education.* New York: Teachers College Press.

Husbands, C., Kitson, A. & Pendry, A. (2003) *Understanding history teaching: teaching and learning about the past in secondary schools,* Maidenhead: Open University Press.

Kagan, D. M. (1992). Implications of research on teacher belief. *Educational Psychologist, 27*(1), 65-90.

Katrin, K. (2016). Sensitive and controversial issues in the classroom: teaching history in a divided society. *Teachers and Teaching, 22*(1), 35-53

Kvale, S. (2011). *Las entrevistas en investigación cualitativa*. Madrid: Morata.

Kovras, I. (2014). *Truth Recovery and Transitional Justice. Deferring human rigths issues.* New York: Routledge.

Legardez, A. & Simonneaux A. (2006). *L'école à l'épreuve de l'actualité. Enseigner les questions vives.* Issy les Moulineaux: ESF éditeur.

Received: 14-03-2017 Accepted: 05-10-2018

Linz, J. (1996). La transición española en perspectiva comparada. In Tusell, J. & Soto, A (eds.). *Historia de la transición. 1975-1986* (pp. 21-45). Madrid: Alianza Editorial.

López-Facal, R. (2003). *La enseñanza de la historia, más allá del nacionalismo*. Madrid: Marcial Pons Historia.

López-Facal, R., Miralles, P., Prats. J. (Dir.) y Gómez, C. (Coord.) (2017). *Enseñanza de la historia y competencias educativas.* Barcelona: Graó.

Low-Beer, A. (1999). *Teaching controversial and sensitive issues in history education for secondary schools.* Strasbourg: Council of Europe.

Maravall, J. M. & Santamaría, J. (1994). El cambio político en España y las perspectivas de la democracia. In O'Donnell, G., Schmitter, P. & Whitehead, L. *Transiciones desde un gobierno autoritario, 1. Europa meridional* (pp. 112-164). Barcelona: Paidós.

McCrum, E. (2013). History teachers' thinking about the nature of their subject. *Teaching and Teacher Education, 35*(1), 73–80.

Monereo, C. & Monte, M. (2011). *Docentes en tránsito. Incidentes críticos en secundaria*. Barcelona: Graó.

Muniesa, B. (2005). *Dictadura y Transición. La España lampedusiana. II: La monarquía parlamentaria.* Barcelona: Universitat de Barcelona.

Ortiz-Heras, M. (2004) Historiografía de la transición. In *La transición a la democracia en España. Historia y fuentes documentales. VI Jornadas de Castilla La Mancha sobre investigación en archivos* (pp. 223-240). Guadalajara: Anabad.

Paulson, J. (2015). "Whether and how?" History education about recent and ongoing conflicto: a review of research. *Journal on Education in Emergencies, 1*(1), 14-47.

Prats, J. & Santacana, J. (2014). Por qué y para qué enseñar Historia. En Prats, J. (coord.). *Didáctica de la Geografía y la Historia* (pp. 13-27). Barcelona: Graó.

Psaltis, C., Carretero, M. y Čehajić-Clanc, S. (2017) (Eds.). *History Education and Conflict Transformation.* Palgrave Macmillan.

Quirosa-Cheyrouze y Muñoz, R. (2009). Los medios de comunicación ante el proceso democratizador en España. En Quirosa-Cheyrouze y Muñoz, R. (Ed.). *Prensa y democracia. Los medios de comunicación en la Transición* (pp. 13–25). Madrid: Biblioteca Nueva.

Rüsen, J. (2005). *History: Narration, Interpretation, Orientation*. Nueva York: Berghahn.

Sáiz-Serrano, J. & López-Facal, R. (2015). Competencias y narrativas históricas: el pensamiento histórico de estudiantes y futuros profesores españoles de educación secundaria. *Revista de Estudios Sociales, 52*, 87-101.

Sánchez, M. (2010). La transición sangrienta. Una historia violenta del proceso democrático en España (1975-1983). Barcelona: Grup Editorial.

Shulman L.S. (1987). Knowledge and teaching: Foundatios of the new reform. *Harvard educational review, 57*(1), 1-21.

Soto, A. (2009). Violencia política y transiciones a la democracia. In Baby, S., Compagnon, O. & González Calleja, E. (Eds.) Violencia y transiciones políticas a finales del siglo XX (pp. 113-129). Madrid: Casa de Velázquez

Toledo, M.I. & Gazmuri, R. (2010). Sobre la enseñanza y el aprendizaje de la dictadura militar y la transición a la democracia en las aulas de secundaria en Santiago de Chile. In *VII Jornades Internacionals de Recerca en Didàctica de les Ciències Socials*. Universidad Autónoma de Barcelona.

Toledo, M. I., Magendzo, A., Gutiérrez, V. & Iglesias, R. (2015). Enseñanza de 'temas controversiales' en la asignatura de historia y ciencias sociales desde la perspectiva de los profesores. *Estudios pedagógicos, 41*(1), 275-292.

Tusell, J. (1994). La transición a la democracia en España como fenómeno de Historia política. *Ayer. Revista de Historia Contemporánea*, *15*, 55-76.

Virta, A. (2001). Student teachers' conceptions of history. *International Journal of Historical Learning, Teaching and Research, 2*(1).

Voet, M. & De Wever, B. (2016). History teachers' conceptions of inquiry-based learning, beliefs about the nature of history, and their relation to the classroom context. *Teaching and Teacher Education, 55*(1); 57-67.

Wineburg, S. (1991) Historical problem solving. A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educatoional Psychology, 83*, 73-87

Yeager E. & Davis O. (1996). Clasroom teachers' thinking about historical texts: an exploratory study. *Theory and research in social education, 24*(2), 146-166.

Received: 14-03-2017 Accepted: 05-10-2018

Yilmaz, K. (2010). Social studies teachers' conceptions of history: Calling on historiography. *Journal of Educational Research, 101*(3), 37–41

Contact address: Rosendo Martínez Rodríguez. Universidad de Valladolid. Facultad de Educación de Segovia. Plaza de la Universidad, 1. 40005, Segovia. E-mail: rosendo.martinez@uva.es.

Back to the drawing board: Can we compare socioeconomic background scales?[1]

Pensémoslo de nuevo: ¿Podemos comparar las escalas de antecedentes socioeconómicos?

DOI: 10.4438/1988-592X-RE-2019-383-400

Andrés Sandoval-Hernandez
University of Bath
David Rutkowski
Indiana University
University of Oslo
Tyler Matta
Pearson
University of Oslo
Daniel Miranda
Pontificia Universidad Católica de Chile, MIDE UC
Centro de Estudios de Conflicto y Cohesión Social

Abstract

Using data from international large-scale assessments (ILSA), we evaluate the issue of country-level model-data consistency of background socio-economic scales, as well as the invariance across countries. To that end, we use data from PISA, TERCE, and TIMSS, as they operationalize socio-economic status somewhat differently. As part of our analysis, we examine whether TERCE, a Latin American study – with measures that are regionally developed – exhibits better psychometric properties than measures that are designed to function across a larger and more diverse number of educational systems. We also examine TIMSS, a trends focused

(1) This paper was supported by the Norwegian Research Council Grant #255246 and by the Center of Social Conflict and Cohesion Studies —COES CONICYT/FONDAP N°°15130009

study – that has historically emphasized consistency and comparison. Finally, we include PISA which has the largest number of participants and has changed and conceptualized a great deal of its background questionnaire depending on the study's major domain and focus. Our findings suggest that none of the socio-economic background scales we analyzed are fully invariant in any of the three studies, and therefore comparisons across countries should be done with caution. The different levels of equivalence reached by each scale in each study and the type of comparisons that can be made given these results (e.g., comparison of average scale scores, comparison of relationships between the tested scales and other variables) are discussed in the full paper.

Key words: measurement invariance, measurement equivalence, TERCE, TIMSS, PISA, multi-group confirmatory factor analysis, socio-economic scales.

Resumen

Utilizando datos de evaluaciones internacionales de gran escala, evaluamos la consistencia y la invarianza de las escalas de antecedentes socioeconómicos de los estudiantes entre los países participantes en estos estudios. Para ello, utilizamos las medidas de antecedentes socioeconómicos desarrolladas por PISA, TERCE y TIMSS, ya que cada estudio operacionaliza esta medida de manera diferente. Como parte de nuestro análisis, examinamos si la escala de TERCE, un estudio latinoamericano – con medidas que fueron desarrolladas con enfoque regional – exhibe mejores propiedades psicométricas que otras medidas que fueron diseñadas para funcionar en un número mayor y más diverso de sistemas educativos. Adicionalmente, examinamos la escala de TIMSS, un estudio enfocado en tendencias – que históricamente ha enfatizado la comparabilidad y consistencia. Finalmente, incluimos también la escala de PISA, que cuenta con el mayor número de participantes y que, en cierta medida, ha cambiado y conceptualizado sus cuestionarios de contexto dependiendo el dominio principal y el foco de cada ciclo del estudio. Nuestros resultados sugieren que ninguna de las escalas de contexto que analizamos son completamente invariantes entre los países que participan en cada estudio, y por lo tanto las comparaciones entre países deben hacerse con precaución. Este artículo discute los niveles de equivalencia alcanzados por cada escala en cada estudio, así como el tipo de comparaciones que se pueden realizar dados estos resultados (e.g. comparación de los promedios nacionales de las escalas, comparación de relaciones o correlaciones entre las escalas evaluadas y otras variables, etc.).

Palabras clave: measurement invariance, measurement equivalence, TERCE, TIMSS, PISA, multi-group confirmatory factor analysis, socio-economic scales.

Received: 27-07-2017 Accepted: 14-09-2018

Introduction

International large-scale assessments (ILSAs) of educational achievement such as the Programme for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMSS), and the Third Regional Comparative and Explanatory Study (TERCE) serve multifold purposes. From system monitoring and benchmarking to providing national participants and researchers with information about what students know and can do. ILSAs offer stakeholders an opportunity for understanding the context and correlates of learning in a number of areas as well as provide important background information on students, teachers, and schools. As national participation in these assessments grows, however, it is becoming more difficult for testing organizations to tailor assessments to meet the needs of a diverse set of participants. For example, in the 2015 PISA cycle, all 34 OECD member countries (representing the largest economies in the world, excepting China) participated in PISA, with the remaining 38 participants (termed *partner* systems) comprised of a heterogeneous mix of economies and cultures, including educational systems such as Tunisia, Peru, Singapore, and Shanghai, China. A similar situation is also faced in TIMSS. Finally, although regional assessments such as TERCE have fewer participants and – ostensibly – less heterogeneity than more global international studies, language, economic and cultural diversity among and within TERCE participants persists. For example, Chile's GDP per capita is over three times that of Bolivia and although the majority of countries share Spanish as a common language many participating countries include indigenous populations with a variety of mother tongues.

Most ILSAs include both a cognitive assessment and a set of background questionnaires. The questionnaires are administered to students and, depending on the assessment, can measure others such as teachers, parents and school leaders. In general, the background questionnaires have two primary uses: (1) to help contextualize the assessed educational system; and (2) to optimize population and sub-population achievement estimation. The benefits of using background data to help estimate achievement are well documented (Mislevy, Beaton, Kaplan, and Sheehan, 1992) and are not the focus of this paper. In fact, potential methodological challenges associated with an amalgamation of participants have been highlighted by a number of researchers, who have pointed especially

to the achievement estimation model and whether comparisons are sensible and valid when systems differ dramatically (Goldstein, 2004; Kreiner and Christensen, 2014; Mazzeo and von Davier, 2009; Oliveri and Ercikan, 2011). Partly in response to these and other criticisms, the PISA project has implemented accommodations, especially targeted toward lower performing participants (e.g., incorporating easier items into the test for countries with low expected performance; OECD, 2012). Recent research has demonstrated that these types of accommodations are a promising way of acknowledging and dealing with the heterogeneity that is necessarily present in cross-cultural research (Rutkowski, Rutkowski and Zhou, 2016). Significant work has been done to ensure comparability of achievement scales across countries (e.g. OECD, 2014; Schulz, Ainley and Frailon, 2011; UNESCO-OREALC, 2016) and across time (e.g. Gaviria and Covadonga, 2007). In contrast, much less effort is spent on designing scales derived from the background questionnaires that can account for vast differences among participants (Rutkowski and Rutkowski, 2010).

Empirically, research has shown that the assumption of equivalent background scales in ILSAs is often violated, leading to compromised comparability (Caro, Sandoval-Hernandez and Lüdtke, 2016; Glas and Jehangir, 2014; Oliveri and von Davier, 2014). As such, the objective of this paper is twofold: First, to demonstrate a method to explore both within-county data consistency and equivalence across countries on background scales. Second, to discuss the results of the application of this method to the socio-economic scales of PISA, TIMSS and TERCE. More specifically, we explore the different levels of equivalence reached by the scales used in each study to measure some form of socio-economic status (SES) and discuss the type of comparisons that can be made given these results (e.g., comparison of average scale scores, comparison of relationships between the tested scales and other variables).

It would not be feasible to evaluate the equivalence of all the background scales of the three studies, for this reason in this paper we focus on the scales developed by the testing organizations to examine some form of SES in three international studies (PISA, TERCE, and TIMSS). We decided to use these scales because, in the studies focused on identifying factors associated with learning outcomes (e.g. school and teacher effectiveness), SES is the control variable that consistently shows a stronger association with educational achievement. Furthermore, there is an important body of literature specifically focused on understanding

Received: 27-07-2017 Accepted: 14-09-2018

the mechanisms by which socio-economic background or family socio-economic status is associated with academic achievement (Buchmann, 2002).

By examining the equivalence of these scales between countries and comparing the findings across studies, we can determine if different assessment designs or approaches result in different degrees of comparability. The three studies were purposely chosen because they represent three different designs of international assessment. TERCE was chosen to represent a regional study – with measures that are regionally developed and with the assumption that test developers were able to focus the scale for a smaller group participants (Treviño, Fraser, Meyer, Morawietz, Inostroza and Naranjo, 2015). TIMSS was chosen to represent a trend focused study – one that has historically emphasized consistency and comparison over changes in societies, constructs, or participants. Finally, we include PISA, which has the largest number of participants and has historically been willing to make significant changes to its background questionnaires (OECD, 2016a).

Analytical framework

Our analytical framework is broadly situated within measurement theory and more specifically within ideas of test theory and design (e.g., van der Linden, 2005; Wilson, 2005). Test theory is focused on how a set of observed responses can map onto a theoretical, unobservable construct. Within ILSAs these observed responses are elicited from a (standardized) test or instrument, which may be defined as "a technique of relating something we observed in the real world (sometimes called manifest or observed) to something we are measuring that only exists as part of a theory (sometimes called latent or unobserved)" (Wilson, p. 4). Instrument design, or the process of developing items that elicit an unobserved theoretical construct, is an iterative process. An underlying assumption in test theory, which governs instrument design, is that the relationship between the theoretical construct and the observed responses to the items that make up the instrument is a causal one (Wilson, 2005). That is, a respondent's level on a particular construct (or constructs in the multidimensional case) causes their responses for a set of items. Because we cannot observe the construct directly, the causal agent is latent and the

Received: 27-07-2017 Accepted: 14-09-2018

measure is left to "infer the underlying construct" allowing the researcher to only assume causality (Wilson, p. 12).

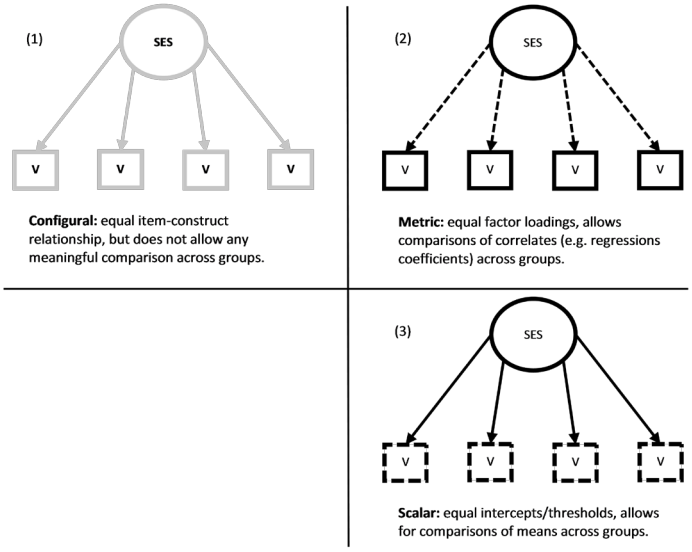
In order to assume a causal relationship between the observed responses and the latent trait, the instrument development requires a rigorous validation process (Cronbach and Meehl, 1955; Messick, 1984)1955; Messick, 1984. In this case, a valid instrument is one where there is ample evidence that suggests the items are measuring the intended theoretical construct for the selected population. The accumulation of the evidence is sometimes referred to as the validation process (Shadish, Cook, and Campbell, 2002). Although the validation process includes multiple steps, one important aspect is to test the correlational structure of the instrument within the intended population of respondents. Assuming the factor structure holds for the given population, the instrument designer then takes additional steps toward validation, such as multitrait-multimethod studies.

When an instrument is intended to be used with multiple populations, as is intended with ILSAs, further validation is required to ensure the instrument operates in the same way across all populations. Brown (2015) outlined four types of group invariance for this purpose: a) equal form, b) equal loadings, c) equal intercepts/ thresholds, and d) equal residual variances (also known as configural, metric, scalar and strict invariance, respectively). The equal form is the most lenient type of invariance and means that the structure of the item-construct relationship is identical across all groups (see quadrant 1 in Figure 1 for a graphical representation). The test of equal loadings builds upon the previous structure and requires that the true score variance in each item is identical across all groups (see quadrant 2 in Figure 1). Next, equal intercepts for continuous items and equal thresholds for discrete items demonstrates that the items have the same locations in the latent space (see quadrant 3 in Figure 1). Finally, equal residual variances, when built upon equal intercepts/thresholds, indicates that all items have the same amount of variance across each group since the loading plus the residual variance equals the total variance. Ensuring measurement invariance indicates that the same construct is being measured in the same way across different groups. Evidence of measurement equivalence does not automatically validate the causal relationship between the construct and respondent; however, an inability to demonstrate equivalence across populations suggests that the assumption of causality between the respondent and

Received: 27-07-2017 Accepted: 14-09-2018

construct does not hold. It is also important to mention that the level of invariance required depends on the objectives of the analysis. Different levels of invariance allow for different types of comparisons. See Figure 1 for a summary of the types of comparisons allowed by the different levels of invariance.

FIGURE 1. Different levels of invariance and types of comparisons allowed in each level



Note: The dotted lines represent the part of the model that is being tested in each invariance level

Methodology

Data

The data for this study has been sourced from the latest cycles of three major ILSAs: TERCE, managed by the UNESCO's Latin American Laboratory for Assessment of the Quality of Education (LLECE); TIMSS, managed by the International Association for the Evaluation of Educational Achievement (IEA), and PISA managed by the Organisation for Economic Cooperation and Development (OECD). All three studies,

TERCE, TIMSS 2015 and PISA 2015 are the most recent international comparative studies that assess student achievement and gather information from a range of educational stakeholders. Specifically, PISA measures student achievement in mathematics, science and reading of 15 years' old students, with a focus on students' ability to apply knowledge in practical contexts and 'everyday life' situations (OECD, 2014, p. 24). In contrast to PISA's focus on practical situations, TIMSS and TERCE are curriculum-based tests and focus on what students had an opportunity to learn in school. TIMSS measures students' achievement in mathematics and science at 4th and 8th grade (Mullis, I.V.S., Martin, M.O., Foy, P., and Hooper, M., 2016), while TERCE measures reading, mathematics and science at 3rd and 6th grades (Treviño, et al., 2015). In this study, we used data from the 72 education systems participating in PISA 2015, from the 44 education systems participating TIMSS 8th grade, and from the 16 education systems that participated in TERCE 6th grade.

For each of the studies, we purposely selected a scale that each testing organisation has constructed and included in their released database as a proxy measure of family background. These scales result from the student questionnaire that is administered to each participant after they complete the cognitive portion of the assessment. In PISA, student's socio-economic status is estimated by the index of economic, social and cultural status (ESCS), which is derived from several variables related to students' family background: parents' education, parents' occupation, a number of home possessions that can be taken as proxies for material wealth and cultural possessions, and the number of books available in the home (OECD, 2016c, p.205). In TIMSS, we used the Home Educational Resources Scale (HERS), which was created based on students' responses concerning the availability of three resources: number of books in the home, highest level of parental education, and number of home study supports (Martin, Mullis, Hooper, Yin, Foy and Palazzo, 2016). In TERCE we used the Family Socioeconomic and Cultural Status Scale (FSCS), which was derived from the following items: parental education, parental occupation, family income, and availability of different home possessions and services (UNESCO-OREALC, 2016). Although the theoretical constructs are not the same across studies, our analyses are intended to examine the extent to which testing organizations are able to create scales that are comparable among the countries that participate in their studies, rather than compare

Received: 27-07-2017 Accepted: 14-09-2018

the same scale across studies. Table 1 shows the set of indicators used in each study to measure socioeconomic background.

TABLE 1. Indicators used in each study to construct a proxy measure of socioeconomic background

Scale / Study	Item	Description
PISA: Index of economic, social and cultural status (ESCS)[2]	1. Highest occupational status of parents (HISEI). 2. Highest educational level of parents (PARED). 3. Home possessions (HOMEPOS).	1. Occupational data for both the student's father and mother were obtained from responses to open-ended questions. The responses were coded to four-digit ISCO codes and then mapped to the international socio-economic index of occupational status. 2. Highest level of education od either parent, recoded onto the following categories: (0) none, (1) primary education, (2) lower secondary, (3), vocational/pre-vocational upper secondary, (4) general upper secondary and/or non-tertiary post-secondary, (5) vocational tertiary and (6) and/or theoretically oriented tertiary and post-graduate. The index corresponds to the higher ISCED level of either parent. 3. Students reported the availability of 16 household items at home, including three country-specific household items and the amount of books at home. Then a summary index of all household and possession items was calculated using IRT modelling with WLEs (logits) for the latent dimensions which were transformed to scales with an average of 0 and a standard deviation of 1 (with equally weighted samples).
TIMSS: Home Educational Resources (HERS)	1. Number of books in the home. (BSBG04). 2. Number of home study supports (BSDG06S). 3. Highest level of education of either parent (BSDGEDUP).	1. Response categories: (1) 0-10, (2) 11-25, (3) 26-100, (4) 101-200, (5) More than 200. 2. Response categories: (1) None, (2) Internet connection or own room, (3) Both. 3. Highest level of education of either parent, recoded onto the following categories: (1) finished primary or some lower-secondary or did not go to school, (2) finished lower-secondary, (3) finished upper-secondary, (4) finished post-secondary education, (5) finished university of higher.

(2) These variables are in turn derived from a set of individual items. See (OECD, 2016c) for more details on the procedure followed to construct this scale.

Received: 27-07-2017 Accepted: 14-09-2018

TERCE: Family Socioeconomic and Cultural Status Scale (FSCS)	1. Highest level of education of the mother (DQFIT09_02). 2. Highest occupational level of the mother (DQFIT11_02). 3. Monthly household income (DQFIT12). 4. Material the floor is made of in the home (DQFIT14). 5. Services in the home (BIENES1). 6. Home possessions (BIENES2). 7. Number of books in the home (DQFIT21).	1. Response categories: (1) none, (2) primary education, (3) more than primary education 2. Response categories: (1) has never worked out of the household, (2) cleaning, maintenance, construction, farmer, etc. (3) sales, operated machines, driver, etc., (4) administrative, owner of smal business, (5) professional, owner of medium/large business, managerial, etc. 3. Income declared recoded into country-income deciles with the following categories: (1) decil 1, (2) decil 2, (3) decil 3, (4) decil 4, (5) decil 5, (6) decil 6 to 10. 4. Response categories: (1) dirt, (2) cement or non-polished wood, (3) tiles or similar, (4) carpet, parquet or polished wood. 5. Students reported the availability of 5 services in the home: drainage, garbage collection, telephone landline, cable TV and internet connection . Then a summary index of all service items was calculated using principal component analysis (PCA). 6. Students reported the amount of the following household items in the home: TV, .radio, PC, refrigerator, washing machine, smart phone, car Then a summary index of all home possessions was calculated using principal component analysis (PCA). 7. Response categories: (1) none, (2) 10 or less, (3) 11-20, (4) 21-30, (5) more than 31

Source: OECD, 2016c; Martin, et al., 2016; UNESCO-OREALC, 2016.

Analytical strategy

Our analytical strategy consisted of two main steps. We first used confirmatory factor analysis (CFA) to test the factor structure of the model used in each study (i.e. TIMSS, PISA and TERCE) to measure some aspect socioeconomic background[3]. One CFA model was fit separately for each country in each study. Then, we used multi-group confirmatory factor analysis (MGCFA) to test for different levels of invariance across countries for each scale in each of the three studies described above. MGCFA (Jöreskog, 1971) is one of the most commonly used techniques to

(3) Even though the TIMSS scale is not a socioeconomic index, the Home Educational Resources Scale is commonly used in IEA's publications as measure to proxy student socioeconomic background. See for example: Martin et al., 2013; Erberber, et al., 2015; Trude and Gustafsson, 2016.

Received: 27-07-2017 Accepted: 14-09-2018

assess measurement invariance (Billiet, 2003). MGCFA is a straightforward extension of CFA that is used to evaluate group differences in means and covariances within a common factor model (Jöreskog, 1971); or as McGrath (2015) puts it, to evaluate overall model fit across multiple groups (education systems in our case).

In the first step, in order to evaluate the goodness of fit for each model in each country, we used four measures: the chi-squared test, comparative fit index (CFI), Tucker-Lewis index (TLI) and root mean square error of approximation (RMSEA). We followed the cut-off points proposed by Rutkowski and Svetina (2014) for the analyses in contexts where the number of groups is large and the sample sizes are large and varied (e.g. ILSA samples): ≤ .10 for RMSEA, ≥ .95 for CFI and TLI. Although the chi-square test is not considered to be useful in this context (Meade et al., 2008; Rutkowski and Svetina; 2014, Cheung and Rensvold, 2002), we also report chi-square statistics in order to analyse if the scales behave as expected across all conditions in that these values are generally larger as more constraints were placed on these models.

It is important to note that for those cases in which the socioeconomic scale is formed with only three indicators, such as in TIMSS and PISA, the one-factor solution is just-identified (i.e. does not have degrees of freedom). As a consequence, the evaluation of fit indexes is not possible because a three-indicator model has a perfect fit. In any case, according to Brown (2015) these "model[s] can still be evaluated in terms of the interpretability and strength of its parameter estimates (e.g., magnitude of factor loadings)" (pp. 71).

In the second step, in order to test for the invariance of the socioeconomic scales across groups (i.e. education systems), MGCFA models were fitted to all groups simultaneously within each study. That is, one MGCFA model was fit to all countries participating in TERCE, a second MGCFA model was fit to all countries participating in TIMSS, and a third MGCFA model was fit to all countries participating in PISA. According to the common practice in the field, we conducted a series of nested tests that proceed from least to most restrictive models. In this way, we started by testing each scale for configural invariance, followed by metric and scalar invariance. Although it is possible to test for strict invariance or equal residual variances (i.e. the fourth level of invariance) in the hierarchy proposed by Brown (2015), scalar invariance is sufficient

for meaningful comparison of latent means across groups (Marsh et al., 2010; Meredith 1993).

We carried out two sets of analyses within this second step. First, in order to examine the performance of MGCFA fit measures, MGCFA models were fit to all countries simultaneously, by study, where the test of configural invariance was followed by the tests of metric and scalar invariance. Following Rutkowski and Svetina (2014), we term this first set of analyses as *overall fit measures.* We evaluate each model (i.e. configural, metric and scalar) using the same criteria presented above. That is, CFI and TLI should be no smaller than .95, and RMSEA should be no larger than .10.

Then, in order to test the plausibility of metric and scalar invariance, we use ΔCFI, ΔTLI, and ΔRMSEA between more and less restrictive models (configural vs. metric, and metric vs. scalar). We term this second set of analyses as *relative fit measures.* Considering the large and varying sample sizes and the relative high number of groups (i.e. educational systems), we use the approach proposed by Rutkowski and Svetina (2014). For the test to show metric invariance, these differences must be ΔCFI ≤ 0.020, ΔTLI ≤ 0.020, and ΔRMSEA 0.030. For the test to show scalar invariance, these differences must be ΔCFI ≤ 0.010, ΔTLI ≤ 0.010, and ΔRMSEA 0.010.

Results

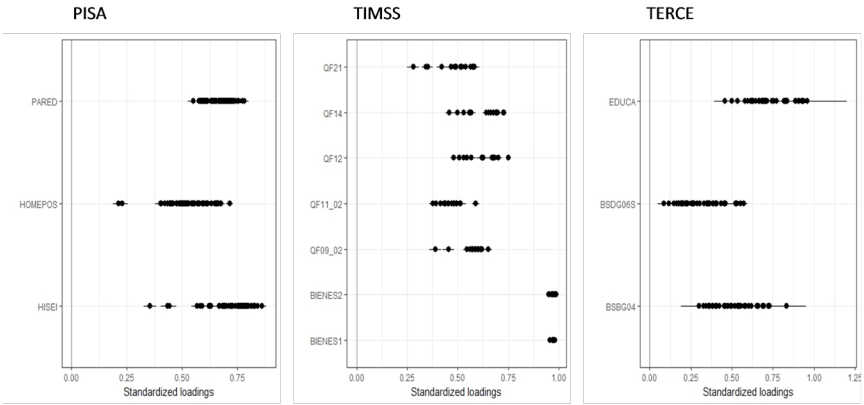
First, we show the general results regarding the extent to which the empirical indicators correspond to the theoretical constructs proposed by each study, tested by CFA procedure for each study/scale and for each country. Second, we show the results of the multi-group analyses and the test for measurement invariance for each study/scale across the education systems participating in each study.

Step 1: Single-country analysis. We start our analysis with separate CFAs for each country in each study. Because the scales for TIMSS and PISA have only three items, there is one unique set of parameters that perfectly fit and reproduce the data (Harrington, 2009). For this reason, instead of presenting a table with the fit indexes (which would include only constant values), we follow the follow the approach proposed by Miranda and Castillo (2018) and present a graph showing the standardised

Received: 27-07-2017 Accepted: 14-09-2018

factor loadings for each item. This allows us to evaluate the models in terms of the magnitude of the factor loadings of each item (Brown, 2015). Figure 1 shows the factor loadings for the PISA scale, Figure 2 for the TIMSS scale, and Figure 3 for the TERCE scale. Even when the model for TERCE is over-identified (df > 0), for consistency purposes, we present the graph with the standardised factor loadings. In Figures 1, 2 and 3, each dot represents the standardised factor loading of each item in one given country, and the horizontal line crossing each dot represents the confidence interval at the 95% level. We marked a vertical line at a 0.5 factor loading as this can be considered the minimum acceptable value for standardised loading in CFA (Hair et al., 2006).

FIGURE 2. Standardised factor loadings for each item composing SES in PISA, TIMSS and TERCE



Received: 27-07-2017 Accepted: 14-09-2018

TABLE 2. Number of countries under/above 0.5 factor loading by indicator/study

Study	Indicator	Factor loading	
		< 0.5	> 0.5
PISA	HISEI	3	65
	PARED	0	68
	HOMPOS	25	43
TIMSS	BSBG04	16	28
	BSDG06S	37	7
	EDUCA	1	43
TERCE	QF09_02	0	16
	QF21	7	9
	QF11_02	12	4
	QF12	1	15
	QF14	1	15
	BIENES1	0	16
	BIENES2	0	16

Figure 2 presents the factor loadings of each indicator used for measuring socioeconomic background in the three studies, and Table 3 shows the number of countries in which the factor loadings of each indicator are below and above 0.5 for each of the studies considered. As can be observed, for PISA, the indicators PARED and HISEI show factor loadings above 0.5 values in most countries (68 and 65, respectively. See Table 2). The indicator HOMEPOS presents 25 countries with factor loadings below 0.5, and even two countries with values under 0.25 (see Table 2). For TIMSS, as can be observed in Figure 2, the indicator EDUCA is the only one that has factor loadings above 0.5 for most countries (43, see Table 2). The other two indicators present higher variations across countries. For instance, the indicator BSBG04 presents 16 countries with factor loadings under 0.5, and the indicator BSDG06S presents factor loadings under 0.5 in 37 countries, with about half of these countries with values under 0.25 (see Table 2). Finally, in TERCE, Figure 2 shows that none of the indicators presents factor loadings under 0.25. The items QF21 and QF11_02, however, present some countries with factor loadings under 0.5 (7 and 12, respectively. See Table 2). Particularly BIENES1

Received: 27-07-2017 Accepted: 14-09-2018

and BIENES2, present factor loadings above 0.5 in all the 16 countries participating in the study (see Table 2).

So far, we have illustrated that the analysed socioeconomic background measures and their configuration across countries show important variations among studies. Our results suggest that among the socioeconomic background scales analysed, the TERCE scales is the one with least variations in its configuration across countries and the one with the best fit, followed by the PISA and TIMSS scales.

Step 2: Multi-group analysis. The test of measurement invariance indicated different levels of invariance for the three analysed studies. In the case of PISA, using the information from the factor loadings, heuristically, it is reasonable to assume that the structure of the scale is similar across countries (see Figure 2). The three indicators had relatively stable estimates across countries, with factor loadings over 0.50 for most indicators in most countries. Only the HOMEPOS index showed some factor loadings under 0.25 (in Qatar and the United Emirates). As can be observed in Table 3, the metric model showed fit indices over the cut-off criteria, while the scalar model showed fit indices under the established criteria (see Table 3). However, the *relative fit* measures indicate that neither metric nor scalar invariance was achieved (see Table 4).

TABLE 3. MGCFA overall fit measures for each level of invariance

	PISA				**TIMSS**				**TERCE**			
Model	X^2	*CFI*	*TLI*	*RMSEA*	X^2	*CFI*	*TLI*	*RMSEA*	X^2	*CFI*	*TLI*	*RMSEA*
Configural	0.000	1.000	1.000	0.000	0.000	1.000	1.000	0.000	4124.600	0.979	0.968	0.070
Metric	5951.480	0.974	0.961	0.077	3195.150	0.863	0.790	0.077	6216.210	0.968	0.965	0.072
Scalar	61932.700	0.728	0.793	0.177	19723.990	0.146	0.672	0.096	16862.100	0.910	0.925	0.107

For TIMSS, the *overall fit* information (e.g. factor loadings) shows that baseline model of configural invariance indicates high dispersion in factor loadings (see Figure 2). Particularly, in only seven countries we found factor loadings for the 0.05, and about one-third of the countries (e.g. Canada, Hungary, Ireland, Italy, Japan, Kuwait) had factor loadings under 0.25. The other two indicators are relatively stable across countries.

The metric model showed *overall* fit indices under the cut-off criteria, and as a consequence, there is no evidence to suggest that metric or scalar invariance was achieved (see Table 3). Similarly, the results of the *relative* fit indices suggest that neither metric nor scalar invariance was achieved (see Table 4).

Finally, the TERCE study showed good *overall* fit indices for the configural and metric models, but fit indices were out of the acceptable range for the scalar model (see Table 3 and Figure 2 for the factor loadings of the configural model). Regarding the *relative* fit measures, the comparison between the configural and metric models provide evidence of metric invariance (see Table 4).

In summary, our analyses resulted in fit indices that did not provide evidence of metric or scalar invariance for the socioeconomic background scales used in TIMSS and PISA, while the TERCE scale showed evidence of metric and scalar invariance.

TABLE 4. MGCFA relative fit measures for each level of invariance

	PISA			**TIMSS**			**TERCE**		
Model	X^2*diff.*	Δ*CFI*	Δ*RMSEA*	X^2*diff.*	Δ*CFI*	Δ*RMSEA*	X^2*diff.*	Δ*CFI*	Δ*RMSEA*
Metric	5951.476	0.026	0.077	3126.102	-0.137	0.077	2086.991	-0.011	0.002
Scalar	44692.380	0.246	0.100	17830.425	-0.717	0.019	7523.642	-0.058	0.035

Discussion

At a basic level, background questionnaires are made of up of multiple instruments, parts of which are intended to measure a hypothetical construct (e.g., socioeconomic status). Because hypothetical constructs cannot be measured directly, answers to select background questionnaire items serve as an indirect indicator of the construct, operationalized through a measurement model. Because constructs are theoretical, unobservable phenomena, the act of verifying or validating the instrument is an extremely important process, a process that falls under modern test theory (e.g., van der Linden, 2005; Wilson 2005). The core idea of

Received: 27-07-2017 Accepted: 14-09-2018

developing and validating such instruments is to begin with a well-defined construct, design a set of items that are assumed to elicit that construct, and then test if the proposed measurement model is consistent with the data generated from those items. When a proposed model does not fit data from one or more items, the items are revised or replaced. In other words, construct development should generally not be a posthoc exercise, where item responses are explored for viable constructs, which are then mapped back onto some theory. At least, it should not be a linear exercise that is finished with the best (but insufficient) attempt to fit the empirical data to a given theory.

Furthermore, only after sufficient evidence suggests the instrument is fitting well within a population can the task of evaluating the suitability of the instrument for cross-population comparisons begin. A common approach to testing measurement invariance across countries is to test the equality of the measurement model's covariance structure, means, and residual variances across countries. By examining the measurement model's degree of equality, we are able to statistically test the assumption of scale comparability. If the assumption holds then there would be statistical evidence that the scales can be compared sensibly. But, as was the case in the current examination, constructs are not always comparable suggesting that the traits being measured are not the same across cultures.

When scales are found to be non-invariant across populations or cultures a number of plausible explanations exist. First, and foremost, it is completely possible that a theorized construct is simply wrong or that it was once correct but changes in society have occurred so that the specified construct is no longer relevant. Of course, if the construct is not relevant the scale should not be reported and the old construct should be abandoned for a new construct. In situations where there exists strong theoretical backing for a universal construct there are other possible reasons that a scale may be found to be non-invariant to include:

1) The construct can be measured but the framework is incorrect;
2) The construct can be measured, the framework is correct, but the indicators are being operationalized incorrectly;
3) The construct can be measured, the framework is correct, but there are no universal indicators.

Received: 27-07-2017 Accepted: 14-09-2018

In terms of the first point, a viable and relatively straightforward treatment is that the framework used to operationalize the construct in question needs to be revised. The second point is a bit more nuanced. Operationalizing constructs incorrectly could happen for a variety of reasons. For example, the framework that stands as the foundation for measuring children's SES should include household income, which is a difficult question to reliably obtain from young children. Thus, other, more indirect indicators of household income must be collected. Potential alternatives could include the number of televisions, bedrooms, or books in the home. Although these might be the most reasonable variables to collect under the circumstance, measures of home possessions may not accurately reflect household income. For example, it is possible that with the dawn of e-readers the number of books in the home no longer represents either wealth or SES. In the third scenario, the construct exists but the indicators needed to measure that construct differ between countries or regions. Again, SES is a useful example to illustrate this point. The majority of academic theory may define SES as a universal construct. In support of the construct, a universal framework might be applied by researchers wishing to measure SES internationally. But regardless of the accepted theory of SES, the indicators that represent the construct may differ by country. For example, a reliable indicator of family wealth in the U.S. might be whether the child has a room of their own or whether they have taken an international vacation. In contrast, a relatively poor indicator would be if the family has a lawnmower. In contrast, a lawnmower is a strong signal of wealth in Hong Kong or Singapore, given the relative lack of land on which to grow grass. The question of universal indicators but not a universal theory remains –is there a set of indicators that can reliably differentiate between well– and poorly-resourced children internationally? Clearly, given the performance of the measures examined here as well as similar research (Caro, Sandoval-Hernandez and Lüdtke, 2016; Rutkowski and Rutkowski, 2017) much work remains to be done.

One possible way forward is to relax the requirement that constructs should be identically defined across measured systems. Although PISA, as one example, has allowed for the inclusion of country-specific wealth items, these are not subsumed under a single latent variable model of SES. Rather, they are treated as observed, country-specific variables that are formed into linear combinations of variables to make up a measure

Received: 27-07-2017 Accepted: 14-09-2018

of socio-cultural status. However, such linear combinations are not latent variables (Bentler, 1982), and have no hypothesized structure. Such approaches are not measuring anything but are merely exercises in data reduction.

It is possible to instead fit latent variable models that adhere to an assumption of partial invariance, whereby unique items and unique item parameters are allowed. Previous research has shown that, although more work needs to be done to operationalize this construct, it is a promising way to improve model-data consistency across countries while maintaining comparability. Importantly, Rutkowski and Rutkowski (2017) concentrated their efforts on the relatively homogeneous Nordic region and showed that more research needs to be done to develop well-functioning country-specific measures. Although this certainly will necessitate meaningful work on the part of participating countries, it will allow for participants to incorporate the local cultural nuances of their local context into internationally comparable scales.

TERCE provides another possible solution. As a regional assessment that focuses on similar language groups, cultures, and economies (when compared to PISA and TIMSS), with more focus TERCE should be able to design and administer questionnaires that are better tailored to a specific population. In the current manuscript, our results indicate that TERCE was able to develop a socioeconomic background scale that is comparable at the metric level which is better than its TIMSS counterpart. Regardless no study had acceptable scalar invariance where latent means can be validly compared across countries. In other words, for TERCE TIMSS and PISA, there is statistical evidence to suggest that the socio-economic background indicator is not cross-culturally comparable. Even worse, in both PISA and TIMSS the scales are not meeting basic quality standards within many participating educational systems. As such, analyses that use mean values of the socioeconomic scales on any of these studies will produce findings that are questionable, at best. These findings have direct policy and research implications. For example, studies that estimate the share of resilient students[4] in a group of countries and then make cross-national comparisons are a classic example of such practice (e.g. OECD, 2011; Erberber, Stephens, Mamedova, Ferguson and Kroeger, 2015). Furthermore, the same can be said about any international comparative

(4) Commonly defined as students with low SES and high academic achievement.

study that uses the PISA index of economic, social and cultural status (ESCS) or the TIMSS' home educational resources scale (HERS) as a control variable in a regression model[5]. Our findings pose a serious threat to the validity of these scales and any future analysis should caution readers to these threats.

Conclusion

Scales from the background questionnaire play an important role in helping to explain educational achievement. In fact, certain scales have taken a life of their own and often operate outside of the achievement results. For example, scales such as bullying, student engagement, and civic engagement are important to policy and interesting in absence of their relationship to achievement. Although, we used SES, a scale common to all three assessments, as an example for this study a similar analysis should be completed for any study that wishes to use scales from ILSAs for cross cultural comparison. Further, as demonstrated in this paper, substantial work needs to be done to improve measures internationally. At the very least, ILSA background scales require a validation process as rigorous as the achievement scales (OECD, 2014). Such a process would go a long way in preventing reporting scales that are not comparable across participating countries.

As hinted at by our results, by embracing a more rigours regional focus to questionnaire development, ILSAs might improve the comparability of certain constructs such as SES. More specifically, improving regional development of questionnaires or further funding the development of regional ILSAs and there questionnaires are two possible ways forward. It could be argued that the IEA's International Civic and Citizenship Study (ICCS), with its regional modules, represents the former and TERCE represents the latter of these possible models. In each case, however, a clear framework that maps directly onto regional specific scales is currently lacking and would need to be fully developed. In the case of larger ILSAs such as PISA and TIMMS, we recommend, at the very

(5) According to our results this type of comparison would be valid when using the TERCE's family socioeconomic and cultural status scale (FSCS) as this scale reached metric invariance (see tables 4 and 5).

Received: 27-07-2017 Accepted: 14-09-2018

least, diversifying the cultural makeup of those stakeholders who oversee the current international frameworks. For example, the expert group that oversaw the PISA 2015 questionnaire framework and instruments committee included eleven members that were mostly from highly developed OECD economies (over half from the U.S. and Germany). The committee composition clearly did not represent the extremely diverse cultural makeup of PISA participants (OECD, 2016b). Finally, in cases where countries or groups of countries find that the framework is miss-specified, members should work with the testing organizations to make adjustments to the framework and scales. If that is not possible, then participants should ask ILSA organizations not to include their country in any scale reporting. This engagement, of course, comes with a cost; however, publishing and using poor scales can be even more costly.

References

Bentler, P. M. (1982). Confirmatory factor analysis via non-iterative estimation. A fast inexpensive method. *Journal of Marketing Research*, *25A*(5), 309-318.

Billiet, J. (2003). Cross-cultural equivalence with structural equation modeling. En J. Harkness, F. Van de Vijver and P. Mohler (Eds.), *Cross-cultural survey methods (*pp. 247-264). NJ: John Wiley and Sons.

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*. New York: Guildford Press.

Buchmann, C. (2002). Measuring family background in international studies of education: Conceptual issues and methodological challenges. En A. C. Porter and A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp. 150–197). Washington, DC: National Academy Press.

Caro, D. H., Sandoval-Hernández, A. and Lüdtke, O. (2016). Cultural, social, and economic capital constructs in international assessments: an evaluation using exploratory structural equation modeling. *School Effectiveness and School Improvement*, *25*(3), 433-450.

Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255. DOI: 10.1207/S15328007SEM0902_5

Cronbach, L. J., and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302.

Erberber, E., Stephens, M., Mamedova, S., Ferguson, S., and Kroeger, T. (2015). Socioeconomically disadvantaged students who are academically successful: Examining academic resilience cross-nationally. *IEA's Policy Brief Series*, No. 5, Amsterdam: IEA. Recuperado de http://www.iea.nl/policy_briefs.html

Gaviria Soto, J. L., Biencinto López, M. C., and Navarro Asencio, E (2007). Invarianza de la estructura de covarianzas de las medidas de rendimiento académico en estudios longitudinales en la transición de Educación Primaria a Secundaria. *Revista de Educación*, 348, 153-173.

Glas, C., and Jehangir, K. (2014). Modeling country-specific differential item functioning. En L. Rutkowski, M. von Davier, and D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis*. Boca Raton, FL: Chapman and Hall / CRC Press.

Goldstein, H. (2004). International comparisons of student attainment: Some issues arising from the PISA study. *Assessment in Education: Principles, Policy and Practice*, *11*(3), 319–330. DOI: 10.1080/0969594042000304618

Harrington, D. (2009). *Confirmatory Factor Analysis*. Oxford: Oxford University Press.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*(4), 408-426.

Kreiner, S., and Christensen, K. B. (2014). Analyses of model fit and robustness. A new look at the PISA scaling model underlying ranking of countries according to reading literacy. *Psychometrika*, *79*(2), 210–231. DOI: 10.1007/s11336-013-9347-z

Martin, M.O., Mullis, I.V.S., Hooper, M., Yin, L., Foy, P. and Palazzo, L. (2016). Creating and Interpreting the TIMSS 2015 Context Questionnaire Scales. En M. O. Martin, I. V. S. Mullis, and M. Hooper (Eds.) *Methods and Procedures in TIMSS 2015*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center / IEA.

Received: 27-07-2017 Accepted: 14-09-2018

Marsh, H.W., Ludtke O., Muthen, B., Asparouhov, T, Morin, A.J.S., et al. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychol. Assess. 22*(3), 471–91.

McGrath, R. E. (2015). Measurement Invariance in Translations of the VIA Inventory of Strengths. *European Journal of Psychological Assessment*. On-line advanced publication. DOI: 10.1027/1015-5759/a000248

Mazzeo, J., and von Davier, M. (2009). Review of the Programme for International Student Assessment (PISA) test design: Recommendations for fostering stability in assessment results. *Education Working Papers EDU/PISA/GB*. Paris: OECD.

Messick, S. (1984). The Psychology of Educational Measurement. *Journal of Educational Measurement*, *21*(3), 215–237.

Miranda, D. and Castillo, J. C. (2018). Measurement model and invariance testing of scales measuring egalitarian values in ICCS 2009. En A. Sandoval-Hernandez, M. M. Isac and D. Miranda (Eds.) *Teaching Tolerance in a Globalized World*. Cham: Springer International Publishing

Mislevy, R. J., Beaton, A. E., Kaplan, B., and Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, *29*(2), 133–161.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika 58*(4), 525–43

Mullis, I.V.S., Martin, M.O., Foy, P., and Hooper, M. (2016). *TIMSS 2015 International Results in Mathematics*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center / IEA.

OECD. (2011). *Against the Odds: Disadvantaged Students who Succeed in School*. Paris: OECD Publishing.

OECD. (2012). *PISA 2009 Technical Report*. Paris: OECD Publishing.

OECD. (2014). *PISA 2012 Technical Report*. Paris: OECD Publishing.

OECD. (2016a). *PISA 2015 Assessment and Analytical Framework.* Paris: OECD Publishing.

OECD. (2016b). PISA 2015 Background questionnaires. Annex A (pp. 129–196). En *PISA 2015 Assessment and Analytical Framework*. Paris: OECD Publishing.

Oliveri, M. E., and Ercikan, K. (2011). Do different approaches to examining construct comparability in multilanguage assessments lead to similar conclusions? *Applied Measurement in Education*, *24*(4), 349–366. DOI: 10.1080/08957347.2011.607063

Oliveri, M. E., and von Davier, M. (2011). Investigation of model fit and score scale comparability in international assessments. *Psychological Test and Assessment Modeling*, *53*(3), 315–333.

Oliveri, M. E., and von Davier, M. (2014). Toward increasing fairness in score scale calibrations employed in international large-scale assessments. *International Journal of Testing*, *14*(1), 1–21. DOI: 10.1080/15305058.2013.825265

Rutkowski, L. and Rutkowski, D. (2010). Getting it "better": The importance of improving background questionnaires in International Large-Scale Assessment. *Journal of Curriculum Studies, 42*(3), 411–430. DOI: 10.1080/00220272.2010.487546

Rutkowski, L. and Rutkowski, D. (2017). Improving the comparability and local usefulness of international assessments: A look back and a way forward. *Scandinavian Journal of Educational Research, 0*(0), 1–14. DOI: 10.1080/00313831.2016.1261044

Rutkowski, L., Rutkowski, D. and Zhou, Y. (2016). Parameter estimation methods and the stability of achievement estimates and system rankings: Another look at the PISA model. *International Journal of Testing, 16*(1), 1–20.

Rutkowski, L., y Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, *74*(1), 31-57.

Schulz, W., Ainley, J. and Fraillon, J. (Eds.). (2011). *ICCS 2009 Technical Report*. Amsterdam: International Association for the Evaluation of Educational Achievement.

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company.

Treviño, E., Fraser, P., Meyer, A., Morawietz, L., Inostroza, P. and Naranjo, E. (2015). *Informe de Resultados TERCE. Factores Asociados*. Santiago: Oficina Regional de Educación para América Latina y el Caribe (OREAL/UNESCO).

UNESCO-OREALC. (2016). *Reporte Técnico. Tercer Estudio Regional Comparativo y Explicativo, TERCE*. Santiago: Oficina Regional de Educación para América Latina y el Caribe (OREAL/UNESCO).

Van Der Linden, W. J. (2005). A Comparison of Item-Selection Methods for Adaptive Tests with Content Constraints. *Journal of Educational Measurement*, *42*(3), 283-302.

Received: 27-07-2017 Accepted: 14-09-2018

Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Contact address: Andrés Sandoval-Hernández. University of Bath, Faculty of Humanities & Social Sciences, Department of Education. University of Bath, Claverton Down, Bath, BA1 6TP, United Kingdom. **E-mail:** A.Sandoval@bath.ac.uk

Received: 27-07-2017 Accepted: 14-09-2018

Education and Cultural Diplomacy in the *Spring of Europe* (1948-1954)[1]

Educación y diplomacia cultural en la *Primavera de Europa* (1948-1954)

DOI: 10.4438/1988-592X-RE-2019-383-401

Luís Domínguez Castro
José Ramón Rodríguez-Lago
Universidad de Vigo

Abstract

The role played by schools and education in the so-called *European spring* has not been sufficiently assessed. Institutions such as the Western European Union, the UNESCO, the European Movement or the Council of Europe have organised a great number of seminars and meetings attended by primary and secondary teachers, education inspectors and intellectuals who discussed how to teach the values of peace and international cooperation and how to introduce the idea of Europe at schools. By applying the theoretical principles of Nye´s soft and relational power and using a qualitative and historiographic methodology, we aim at analysing those meetings documents in order to identify the resources are used by each state in their fight for cultural hegemony in education, and to lay the foundations for Europe to be taught at schools. Thus, Modern Languages and History are the most studied disciplines. It is also stated that the teaching of a foreign language is necessary in primary school curricula; English leads the way as schooling language and French fights for keeping its hegemony resorting to a highly active cultural diplomacy. A general European History book won´t

[1] This study was co-funded by the Spanish *Ministerio de Economía y Competitividad* (project HAR2015-64429-C2-1-P) and the Education, Audiovisual and Culture Executive Agency of the European Commission (Jean Monnet project, 574846-EPP-1-2016-1-ES-EPPJMO-PROJECT). The Spanish version of the article has been translated into English by Macarena Muradás Sanromán, *Universidad de Vigo*.

be warmly received, but to lessen the excesses of national myths in text books is said to be indispensable. Finally, it is considered that recent history must own a place in the curricula so as to avoid the risks of its possible manipulation by European younger generations.

Keywords: school, soft power, relational power, education, cultural diplomacy.

Resumen

El papel que jugó la escuela y la educación en los debates durante la denominada "primavera de Europa" no ha sido suficientemente valorado. Instituciones como la Unión Occidental, la UNESCO, el Movimiento Europeo o el Consejo de Europa organizaron un buen número de seminarios y encuentros de maestros, profesores, inspectores e intelectuales que reflexionaron sobre cómo formar a la ciudadanía en un marco de paz y cooperación internacional y cómo introducir la idea de Europa desde la escuela. Aplicando los principios teóricos del poder blando y relacional de Nye y utilizando una metodología cualitativa e historiográfica, analizamos la documentación conservada de las reuniones para ver qué recursos emplea cada Estado en la lucha por la hegemonía cultural en el ámbito educativo y para acordar los principios inspiradores de la idea de Europa a transmitir desde la escuela. Las lenguas modernas y la historia van a ser las disciplinas que más atención reciban. Se afirma que es necesaria la presencia de una lengua extranjera en el currículo de primaria. El inglés avanza como lengua escolar y el francés se resiste a perder su hegemonía a través de una diplomacia cultural muy activa. Se rechaza elaborar un manual de historia general de Europa, pero se insiste en la necesidad de mitigar los excesos de la mitología nacional en sus páginas. Se considera por último que la historia reciente, debe tener cabida en el currículo escolar, para sofocar los riesgos de su posible manipulación entre las jóvenes generaciones de europeos.

Palabras clave: escuela, poder blando, poder relacional, educación, diplomacia cultural.

Introduction: theoretical framework

Post-war years immediately after World War II were dramatically influenced by the Cold War, a period of confrontation between two blocs: those led by the USA and the USSR. In western Europe, this conflict

was visible in the cultural field, where they did not skimp on resources, both financial and organisational. The present article is framed in the so-called *Cultural Cold War* (Saunders, 1999), a term which refers to the fight, in all spheres, against the influence of the communist culture in a section of the western *intelligentsia* (Judt, 2007). Schooling has not received any attention in Saunders' book or in more recent monographs (Scott-Smith & Krabbendam, 2004), focused on *high culture* –plastic arts, classical music, literature– and on *mass culture* – visual arts and media–. Therefore, we intend to contribute to fill this gap by revealing the importance of educational debate in four organisations which played a very relevant role during this period: the Western Union (WU), UNESCO, the Council of Europe and the European Movement.

The late 40s and early 50s coincide with the *Spring of Europe* (Bossuat, 1994 & 2005). During this period, culture gained a certain prominence. Nevertheless, other key issues such as the success of economic integration organisations and the failure on the part of political and military cooperation institutions have overshadowed cultural debate in those days, including debate about school and its potential to *socialise a society* committed to international peace and harmony (Gavari Starkie, 2015). Bossuat ends up stating that the *Spring of Europe* comprises the period between 1945 and 1949 because that is the year when customs union –promoted by the OEEC– fails, the opportunity to create a Council of Europe at a supranational level is missed and NATO is the perfect example of dependence on the USA. However, we have decided to adopt a different time frame, since we situate this period between 1948 (when we witnessed the consolidation of the diverse movements in favour of European unity within the European Movement and the birth of organisations such as the WU or the OEEC) and 1954 (when the failure of the European Defence Community awoke those who dreamed of the United States of Europe and made them get back to reality).

An essential theoretical term that will be systematically used in the present article is *relational power* (Nye, 2004 & 2011), which refers to the capacity of turning the potential related to the ownership of resources into real power. Transnational agents –such as the European Movement, the WU, UNESCO and the Council of Europe– played a key role in this new scenario of international relations, without neglecting that performed by the States as organisations with the capacity of creating and maintaining trust networks which would enable to work together in order to reach

common goals. Relational power is always exerted *with* others and not *over* others; therefore, it can be stated that its influence is based on a deft use of *soft power* and *intelligent power* to establish the preferences of others in line with those of the State or the agent that makes use of them (Nye, 1990 & 2004). Given that three of the organisations with which we are working are intergovernmental, it should be noted that the States have three different intangible resources at their disposal in order to help them exert soft power: their culture, their political values and their foreign policy. However, as pointed out by Nye (2004), public diplomacy is only effective when our culture is attractive to others, when our political values are respected by the rest because they are observed in our country of origin and when our foreign policy enjoys legitimacy and moral authority. Public diplomacy can be defined as the group of actions taken by governments in order to address other countries' citizens with the intention of getting control over their public agendas, by modifying their preferences thanks to the influence they have on their expectations about what is legitimate and possible. One of the variants of this new diplomacy, as opposed to the traditional one, is *cultural diplomacy*, which is intended to spread the cultural expressions of a given society, with an essential role being played by education.

Those intellectuals who promoted European unity believed that the educational system had to enable the identification of elites capable of spreading and putting the idea of Europe into practice (*Mouvement Européen*, 1949). A considerable part of these intellectuals belonged, in terms of age and political culture, to the generations who defended that education should only be reserved to elites, in charge of transforming the society from the top without facing those dangers inherent to all changes made from its basis. Nevertheless, the electoral emergence of both British Trade Unions and continental Christian democracy, as well as the consolidation of the Scandinavian social democracy, at a time in which intellectuals' interest in Communism is at its peak, make it necessary to pay attention to mass education, in particular to the first stages of education and access to Secondary. This period of time is also greatly influenced by the introduction of comprehensive schools in Sweden, after a few experiences in other Nordic countries as well as in the United Kingdom (García Pedraza, 2015).

On another note, school was one of the most important nationalising agents –along with the Army– in Europe in the second half of the 19th

century and the first half of the 20th. Therefore, it is not surprising to find that Europeanists regard school as a key factor in the triumph of the new doctrine of unity which was taught from a range of pulpits on both sides of the Atlantic. Nevertheless, when it comes to build a European identity and to favour an atmosphere of agreement and understanding at an international level, not all the subjects included in the school curriculum are equally important. In fact, Modern Languages and History are those which receive more attention; cultural diplomacy of western countries has to face not only the great influence of the Communist universe, but also the effects of decolonisation and the emergence of the so-called *Third World*. France had accumulated a great deal of experience in the use of cultural diplomacy (Haigh, 1974). In the present paper we intend to find out to what extent this is equally applicable to the educational field, what other countries came into scene and what resources –soft, intelligent and relational power– were put into play by each of them. In order to do so, with the aim of analysing the results of the different seminars, symposiums and conferences which were organised by the European Movement, the WU, UNESCO and the Council of Europe, we will resort to a qualitative research methodology (Flick, 2004) which we will combine with the theoretical tools provided by Nye, adapting them to a scenario prior to his postulates, such as that of the *Spring of Europe*[2].

The fight for cultural hegemony in the educational field

Just like Germany, France had been displaying an active culture diplomacy through education, art and archaeology since the end of the 19th century. The USA joined this initiative in the 1930s, especially in Latin America, in an attempt to oust French as the dominant language at American schools, universities and colleges by replacing it with Spanish (Delgado Gómez-Escalonilla, 2015). During the post-war period we witnessed the dispute over cultural hegemony in Europe between France, heir to a privileged position now under threat, and the United Kingdom, bursting with pride as a result of the worldwide expansion of its language, in large

(2) The documentary sources we have consulted are the Archives Historiques de l'Union Européenne (AHUE) at the European University Institute (Florence) and the online archives of both the Council of Europe and UNESCO.

Received: 09-11-2017 Accepted: 14-09-2018

part thanks to the leadership of the USA in the Western Bloc. In fact, Jean-Paul Sartre (1949, p.1) claimed that "*les idées culturelles, tout à fait indépendamment de leur valeur interne, ont un potentiel de diffusion qui dépend de l´importance économique ou militaire du pays considéré… Si nous voulons que la culture française reste, il faut qu´elle soit intégrée aux cadres d´une grande culture européenne*". Countries such as France, Germany, Italy, Netherlands, Sweden, Luxembourg, Greece or Portugal are referred to as members of the European cultural community in this five-page document; in contrast, there is not a single reference to Great Britain. Sartre's words fit in with Nye's notion of *intelligent power*, as it combines aspects of both *hard power* –militia, economy– and *soft power* –culture– and prioritises action based on a relational power which enables to gain support from other countries against the British development, to the benefit of the French interests.

Education becomes one of the fields where the battle for cultural hegemony takes place. Thus, the Cultural Committee of the WU organised three sets of courses, each of which was aimed at a different target group: education inspectors, teachers or public workers. The sequence was always the same. The first courses were held in Great Britain in 1949; the second set took place in France the following year. The most important one for the purposes of this study, the course for teachers, started in 1949 in Ashridge (England), continued in Sèvres (France) in 1950 and concluded in Oosterbeek (the Netherlands) in 1951. All these courses resulted in the publication of the pamphlet *The Civilisation of Western Europe and the School*, which, written by French educator Pierre Joulia and edited in English, French and Dutch, contained a number of suggestions aimed at helping teachers deal with this topic in their classrooms (Haigh, 1974).

A few days after the creation of the WU, Ifor Evans –at the time rector of Queen Mary College and former Head of Education at the British Council (1940-1944)– writes a letter to Kenneth Lindsay –an independent deputy and Minister of Education (1937-1940)–, in which he gives his opinion about the role that should be played by culture, most especially by education, in the European unity process. In fact, Evans (1948) considers that the primary school stage is the perfect time to introduce intercultural knowledge:

Received: 09-11-2017 Accepted: 14-09-2018

"I should like to see prepared in the first place a book for children who are beginning to read which would have the names of the different countries of the Western Union with some of their customs and stories, so that at the age of ten or eleven boys and girls would be as familiar with the names of the members of the Western Union as they are with the words English, Irish, Scots and Welsh. The book I have in mind would be a standard book, but it would be translated into each of the languages of the Western Union and attempts would be made to get the Ministries of Education, Religious Education bodies, etc., to use it."

In other words, this was an attempt to export the *Union Jack* model to the continent; indeed, the United Europe Movement, supported by Churchill, had British and French divisions, in line with the Franco-British Union which the Tory leader himself had proposed in June 1940. The three Benelux governments had spent the war years exiled in London, where intense cultural diplomacy had been exerted and where the British relational power had played a prominent role by means of the Conference of Allied Ministers of Education (CAME).

As for the secondary education stage, Evans (1948) suggests to adopt a different strategy:

"There would be the problem of affecting children at a somewhat more advanced school age and here I would suggest the preparation of two volumes. One would be a very simple Western Union handbook which would give information on the nature of the constitutions, size, government, etc., of each of the countries in the Western Union. It would give something about their resources and the ways in which they could help each other. It would tell of their school life and their Universities and have some reflections upon their cultural achievements in the past. Once again the book would be a standard book, but it would be translated into each of the languages of the Western Union. Its contents would be agreed contents, though I think once again as we are free nations that a certain amount of compromise and variation should be allowed. The other book I would have in mind would be a little more difficult to prepare. It would be a selection from some of the literature of each of the countries of the Western Union. I think it might have to be confined in the first instance to only some ten extracts, verse

Received: 09-11-2017 Accepted: 14-09-2018

or prose, and they would have to be translated as would be the other projects I have outlined...”

It should be noted that secondary education was a selective tool to consolidate elites in the five countries which made up the Western Union. This explains the intention of creating a community to belong to and of instilling cooperation to face an uncertain future due to the presence of external threat. Surprisingly, History (the most remarkable school subject when it comes to building identities) is not mentioned by Evans except for a reference to the cultural milestones achieved by each country in the past. However, the thinker does point out the importance of another essential tool to create identities and stereotypes: literature (López Facal, 2010).

In parallel with this early interest in education on the part of Great Britain as an essential basis on which to build a European narrative, the French division of the European Movement submits a report about the role of education, through Jean Bayet –holder of a chair in Latin language & literature in the Sorbonne and former Head of Education during De Gaulle’s government in 1944– at the European Cultural Conference in Lausanne (1949). In this report, which will serve as a basis for the conclusions of the Conference with the aim of offering a European education at all levels, Bayet resorts to clear elements of soft power in order to justify the right of the French to set the preferences which will be assumed by the representatives of the other countries: “*le rayonnement international permanent de plusieurs des Universités françaises, la forte organisation de l´Enseignement d´État en France, son caractère non-confessionnel, les* études *préliminaires faites à partir de 1945, en vue d´une réforme générale et profondément démocratique de l´enseignement*.” (Bayet, 1949, p. 1).

Besides insisting on the prestige of the French public educational system, the Bayet report claims, and the Conference assumes, that in the final years of primary education it is necessary to introduce the teaching of a living foreign language, paying special attention to oral skills and vocabulary acquisition. This suggestion originates in France’s decision of using the Direct Method for teaching living languages. This method, available since 1902 and inspired by Bismarck’s Germany, gave priority to active oral skills over the traditional written translations inherited from the teaching of Latin and Greek (Puren, 2012).

Received: 09-11-2017 Accepted: 14-09-2018

The prestige of this active language teaching method used in France not only left its mark on the European Cultural Conference held in Lausanne, but also on the Modern Language seminar which, organised by UNESCO as a tool for promoting international understanding, was held in Nuwara Eliya (Sri Lanka) from the 3rd to the 28th of August 1953. In fact, the document submitted by the French division, which bet on an active language teaching method, served as a basis for the preparation of the seminar (Morales Gil, 2007). Nevertheless, Americans counterattacked by defending their Audio-lingual Method (also known as the Army Method), resorting to prestigious Ogden's Basic English Handbook and highlighting the fact that nothing similar existed in the other languages. The Israeli and Australian experts who attended the seminar corroborated the success of this method when applied to the quick integration of migrants into society (UNESCO, 1954). The result of this debate was a series of agreed resolutions which combined both methods.

Once the Council of Europe decided that it was necessary to introduce foreign language learning at school, including primary, the dispute moved on to the languages which should be taught. In 1952, France submitted a proposal to create a European language community based on French-English bilingualism, claiming that both languages were widespread and asking for the inclusion of one of them in primary education in those countries where these languages were not spoken as mother tongues. By doing this, students would have the opportunity to learn another language –either German or Italian– at the secondary education stage. The German representative opposed this proposal, by claiming that a trilingual education system should be used, given the importance of German, spoken by 100 million Europeans from a range of countries in Central and Eastern Europe, not to mention the fact that it was the language of Luther, Goethe and Heine (Council of Europe, 1952b). The fight for supremacy between French and English had undiplomatic episodes at the Council of Europe's Council of Ministers, such as when the Committee of Cultural Experts amended the minutes of one of its subcommittees, which said: "*Au tours de ces dernières années, un grand nombre d´adultes de certains pays ont appris l´anglais qu´ils n´avaient pas appris à l´école*" (Council of Europe, 1952a, p. 13), by translating the abovementioned sentence into English in a much more imprecise way: "*During the last few years, great numbers of adults in some countries*

have learned languages which they never learnt at school" (Council of Europe, 1952a, p. 4), which can be considered a clear example of this soft power battle. Eventually, the Council decided to stand aside and leave the decision in the hands of each member State, by arguing that democratic systems do not allow impositions of this nature and that it might be more useful to speak the neighbour's language than an internationally widespread one (Council of Europe, 1952a).

UNESCO, for its part, also opted not to make any recommendations on what specific languages should be chosen, especially having into account the remaining vestiges of cultural imperialism which, according to Asian experts, rendered the students' empathy with certain western languages more difficult. Moreover, it should be noted that, for Asians, learning such languages would mean a break with their cultural continuum, which was connected to Sanskrit, Pali and Persian, but not with Latin or Greek (UNESCO, 1954).

While France and Great Britain/the USA play the leading roles in the dispute over linguistic dominance, more countries are involved in the fight when it comes to History. In fact, Scandinavians will reclaim their right to be considered pioneers in creating bilateral commissions aimed at dealing with the presence of national prejudice in school books, which can be regarded as a clear example of relational power on the part of the Nordic Council. Indeed, several committees of Danish, Norwegian and Swedish historians had been created as early as the 1920s in order to carry out this task. The Norwegian committee, for instance, could confirm that primary school books described how Trondheim's old cathedral had been used as a stable to keep the horses of the Swedish invaders (1563-1570). This fact becomes even more relevant if we have into account that the remains of Norway's patron saint, Olaf, rest in the cathedral, and that the construction was rebuilt as a national sanctuary in the 19th century. Even if we assume that these facts were true, it should be noted that, at the time, the cathedral was abandoned and in ruins (Vigander, 1950), important information which was not mentioned in the books in order to cause a greater impact on children's imagination. This experience enabled the UNESCO seminar (held in Brussels from the 12th of July to the 23rd of August 1950 with the aim of improving textbooks, especially History books) to support the Nordic model, which included mutual revision by means of bilateral commissions integrating authors, publishers, teachers and academics (UNESCO, 1950). Moreover, this experience resulted in

Received: 09-11-2017 Accepted: 14-09-2018

the appointment of Norwegian Haakon Vigander as the chairman of the Symposium on the idea of Europe in History teaching, held from the 4th to the 12th of August 1953 in Calw (in the Black Forest, Germany) under the auspices of the Council of Europe.

Like in the linguistic field, Germany did not stand aside. Rather the opposite, its cultural diplomacy played an important role in History teaching. In fact, the election of Calw to host the Council of Europe's Symposium had a lot to do with the reconstruction process of the new federal state –still not sovereign–, very concerned with textbook revision. No wonder Germany took part in eight of the thirteen bilateral commissions which were created for textbook revision. Furthermore, through the Internationales Schulbuchinstitut Braunschweig –the current headquarters of the Georg-Eckert-Institut für internationale Schulbuchforschung–, the country threw itself into the organisation of a series of conferences on the topic which were held on an annual basis between 1951 and 1963. In addition, the symposium of Calw was directly supported by the State Secretary of the Foreign Office, Walter Hallstein, who would later become the first president of the European Commission. Significantly, the three presentations on a European concept of History were also assigned to a German expert –Middle Ages–, a French one –Early Modern Period– and a British one –Late Modern Period– (Council of Europe, 1953b).

At UNESCO, just like it happened with languages, History teaching also generated a lively debate around western cultural supremacy. The seminars held at Brussels (1950) and Sèvres (1951) confirmed the need to de-Europeanise Universal History textbooks. The different arguments brought forward in favour of incorporating other civilisations are an example of intelligent power (UNESCO, 1951b, p.8): "*Non-European peoples must be given a place which is proportionate to their world importance and their contributions to civilization. Nor should this importance be measured in terms of territorial size [...] military strength or wealth alone; contributions to humanity´s common heritage must be also taken into account.*". Nevertheless, this criterion eventually results in a clear commitment to increase the presence of the Indian and Chinese civilisations, but not of others, which is a clear indication of the poor relational power of the Latin American nations in UNESCO despite the fact that they represented a third of its member countries at the time.

Received: 09-11-2017 Accepted: 14-09-2018

Schools as a Europeanisation factor

In his report for the European Cultural Conference, Jean Bayet (1949) claims that generating and spreading an "*esprit européen*" is not a task for schools, as this idea had actually been backed not by universities, but by the relationships established between teachers and students who had come to them from all over the world. Bayet also insists on the need to promote exchange as a relational power tool. After the moral and vital disaster in which resulted war and post-war years, the leaders of the Conference believed that no political or social reorganisation could build a sound basis without a radical restructuring of education and teaching which would define the education of young people in all European countries, which had encountered difficulty in the previous centuries because of the loss of a common cultivated language, the autonomy and wealth of national cultures, and the growing importance of utilitarian perceptions of culture and education as opposed to those based on spiritual improvement without an immediate practical aim.

The strong Federalist influence of the main organisers of the Conference –Rougemont, Silva, Madariaga– represents a strong commitment to mitigate Nationalisms, so present in school education: "*les éducateurs de tous pays soient préparés à assumer la tâche primordiale de former les nouvelles générations dans un esprit non seulement national mais aussi européen et largement humaniste*" (Mouvement Européen, 1949, p. 15). However, the Bayet report (1949, p. 2) warned about the inconvenience of a premature and hasty control on the part of an aspiration of uncertain appearance (like the European) over the solid contributions made by national cultures:

> "*Ce n´est qu´à force de siècles, que les provinces françaises se sont groupées en Nation sans perdre trop de leur potentiel propre; une Europe fondée très vite sur une "Koiné" indistincte et sans nuances aurait trop perdu de sa substance et de ses énergies pour être autre chose qu´une fragile illusion de surface déguisant le néant d´une ruine*."

Rougemont (1949), aware of this, insists that there is no intention at all to replace twenty local nationalisms with a single European nationalism. From its very beginnings, Europe has been open to the world. In fact,

Received: 09-11-2017 Accepted: 14-09-2018

it has always understood its civilisation as a combination of universal values. The objective was, therefore, to get future teachers to spread their humanist message like wildfire in a society which was more and more willing to receive it, just like it had once happened with extreme nationalisms. The final declaration of the Conference claimed that national governments should finally admit that investing on education ensured a longer-lasting containment and resistance power than military expenditure. This statement makes even more sense if we take into account that the above-mentioned declaration starts by assuring that "*le problème est simple: il faut faire l´Europe ou il faut faire la guerre*" (Mouvement Européen, 1949, p. 18).

History, especially when it comes to textbooks, becomes one of those disciplines which we can consider key not just in terms of construction and legitimation of national identities, but also with regard to the dissemination of the idea of Europe. The Conference suggests that textbooks should be revised by a European Inter-university Council with the authority to recommend any amendments which may be considered necessary in order to ensure that such books are inspired by a Europeanist and pluralist spirit.

In fact, a concern for the contents of History textbooks had been a part of the post-war scenario since the beginning. UNESCO, since its foundation, pursued the goal of establishing a truly Universal History based on a new approach which, in the case of Europe, could break the monopoly of its Eurocentrism. Indeed, the institution organised two seminars which specifically dealt with History and textbook revision. The first one was held in Brussels (1950), whereas the second took place in Sèvres (1951). The latter showed the study of 24 European textbooks for students aged 12 to 15, whose contents (related to National, European and Universal History) were analysed with the aid of the following table (UNESCO, 1951a, p. 23)[3]:

(3) It should be noted that data about England (obtained after the analysis of 5 textbooks) might not necessarily represent the reality, given the great number of different textbooks which are used in the English educational system. As for France (4 textbooks), the available information should be considered an approximate estimate, in addition to the fact that such information does not distinguish between European History and Universal History. The source does not specify either what the term "National History" means in the English and the Scottish contexts.

TABLE 1. Geographical distribution of History textbook contents

COUNTRY	NATIONAL HST.	EUROPEAN HST.	UNIVERSAL HST.
Switzerland	67.5%	32%	1,5%
Germany (FRG)	32%	53%	15%
England	21%	33%	46%
Belgium	18%	66%	16%
Scotland	51%	16%	33%
France	45%	55%	
Greece	68%	16%	16%

Despite the partially random nature of the sample, this table facilitates reflection on interesting issues. Germany was going through a denazification process, so only a third of the contents cover National History, whereas more than half deal with European History. The case of Belgium is even clearer, with two thirds of the contents assigned to European History, even though it should be noted that only National History was taught in primary schools (Council of Europe, 1953). Greece and Switzerland, for their part, illustrate the most nationalist sector, with a greater Europeanist tendency in the second case. Finally, let us analyse France and England. What happened in the former country is in line with the Bayet quote we presented earlier, which focused on how inopportune it was to accelerate Europeanisation at schools, regardless of the fact that things have actually evolved in that direction, as proposed by Sartre. Finally, although this data should be interpreted with caution, we can affirm that England showed a greater interest in Europe at the time than nowadays, according to recent surveys (Gómez Carrasco & Chapman, 2017).

If UNESCO worked for the improvement of History textbooks with a view to the coexistence of nations and cultures, the Council of Europe followed suit with the aim of consolidating and legitimating the idea of Europe among the new generations. In fact, there is no shortage of authors who talk about a *Eurohistory* inspired by the European Movement and supported by the Council of Europe with a series of scientific meetings on this topic (Davies, 1996). The Council held its first symposium in Calw, from the 4th to the 12th of August 1953, with the objective of

Received: 09-11-2017 Accepted: 14-09-2018

discussing the presence of the idea of Europe in History teaching. The symposium was divided into two parts. The first of them was devoted to the introduction and discussion of the different national reports which had been previously prepared by each delegation to show the results of the revision of History textbooks and teaching in each country, while the second part of the symposium was focused on the analysis of the components of a European concept of History during three different epochs: Middle Ages, Early Modern Period and Late Modern Period. In order to do so, two questions were posed: from among all the general principles that intervene in the construction of a European historical awareness, which ones should be taught in the European countries? how should we tackle any problems which may hinder the construction of a common European historical awareness? (Council of Europe, 1953a). However, as pointed out by Valsecchi, "*Le Symposium n´a pas à élaborer une nouvelle interprétation de l´histoire européenne; il doit […] procéder à une révision politique et logique des manuels scolaires qui contribuera à faire disparaître les égoïsmes nationaux […] il ne convient pas de rechercher dans le passé une Europe qui n´existait pas*." (Council of Europe, 1953a, p. 23-24)

Hübinger, the expert in charge of presenting the report on the Middle Ages, stated that the idea of Europe had arisen in the 15th and 16th centuries. There was not one medieval period in Europe, but several. Nevertheless, certain elements will be revisited when the idea of Europe takes shape, such as Christian, classical and German heritage, the imprint of feudalism, Roman law, University and religious orders. From their perspective, the survival of Roman imperialist ideas was an obstacle rather than a point of agreement, since each emperor claimed to be the legitimate heir to the throne. Moreover, unity between eastern and western cultures was not maintained either, as demonstrated by the fact that Aristotle's doctrine was transmitted through the Arabs, not via Byzantium. Despite the relevance of the crusades, there was not union against Islam in Europe until the arrival of the Turkish threat. The Greek representative, Zakytinos, denied that feudalism had been one of the keys to European awareness, since this phenomenon had exclusively taken place in central Carolingian Europe, which means that a considerable part of Italy and Spain, Scandinavia, the Byzantine Empire and the Slavic countries would be excluded. Neither did he accept the cultural rupture between the East and the West, as, in Paul Valéry's words, "*Partout où*

les noms d´Homère et de Socrate, d´Alexandre et de Platon, d´Aristote, de St. Paul ont une autorité indiscutible –là est l´Europe" (Council of Europe, 1953a, p.18). In reality, those soft power resources employed by the German and the Greek spokespeople are trying to legitimise an idea of Europe which favours their regional interests —East and West— with regard to the construction of this new transnational identity.

In the Early Modern Period it is possible to find certain principles of the European spirit, apparently contradictory, which must be included in the school curriculum. For instance, the protection of human rights is a core pillar of the idea of Europe, which clearly evolved from natural law —both Catholic and Calvinist—, to the English Bill of Rights and the Declaration of the Rights of the Man and of the Citizen created during the French Revolution. However, the opposite *raison d'état* is also a pillar of European identity, just like the dynastic principle and the democratic principle. The creation of States and the emergence of nationalism do not prevent the simultaneous birth of a certain notion of European unity. Universalism is another principle of the European culture which should be taught at school, not as an example of supremacy on the part of the western civilisation, but as a clear sign of Europe's inclusion in Universal History, which adds value to the other cultures.

In the Late Modern Period, freedom is the core principle of the idea of Europe. Moreover, there was some debate about whether Contemporary History and the most recent events should be present in the school curriculum. Finally, it was agreed that this period should not be excluded; otherwise, young people would approach it through political propaganda and through the press. In other words, they would lack the necessary critical apparatus which only school could offer them. Finally, what the experts gathered in Calw discarded was the proposal of recommending the Council of Europe to sponsor the creation of a general European History textbook, not only because this action would be doomed to failure —as the UNESCO experience had proven—, but also because it would go against certain traditions in matters of education, such as the British.

Conclusions

Post-Second World War Europe desires —just like everywhere else in the Globe— to build international relationships which can guarantee

Received: 09-11-2017 Accepted: 14-09-2018

a peaceful future for all nations. Western Europe also hopes that such future is built on a series of values which reinforce democracy and freedom, while scaring away the ghost of communist domination. In this scenario, education played a key role. The *Spring of Europe*, which takes place between The Hague Congress (May 1948) and the failure of the European Defence Community (August 1954), is witness to the great effervescence of initiatives and debates (under the auspices of the Western Union, the European Movement, the Council of Europe and UNESCO) in which school is present. There is a common denominator: the need to transmit values and behaviours which favour cooperation between citizens from different countries and also the need to mitigate, to the extent possible, the damage caused by an educational system biased by nationalist paradigms and obsolete didactic approaches.

Modern Languages and History are two disciplines which merit special attention. In the first case, it is vital to speak different languages because this enables us to understand others better and to favour exchange with the aim of ensuring and consolidating international cooperation without having to choose a specific first foreign language. As for the second case, the teaching of History is known to have strong points such as textbook revision —intended to purge them of nationalist mythologies, in conflict with a new era characterised by long-awaited universal harmony— and the construction of a solid basis on which it is possible to build the inspiring, emblematic principles of the idea of Europe.

The major European powers make use of the many education conferences and seminars held at the time to exert their cultural diplomacy and bring into play their relational, soft and intelligent power resources. France and Great Britain lead the fight for supremacy not only in the cultural field, but also with regard to the organisational and relational design of the new Europe which is emerging. France had a head start thanks to its experience, with important achievements such as the creation of the Écoles *Françaises* in Athens (1873) and Rome (1875), the *Alliance Française* (1883) and Casa Velázquez (1920). As for Great Britain, despite starting later, with achievements such as the creation of the British Council (1934), it was able to take full advantage of its status (as seat of the European governments exiled during the Second World War) to spread its intelligent power —cultural prestige and strong military power— and relational power with regard to COMECON as well as to the organisational design of the European Movement, the Western

Received: 09-11-2017 Accepted: 14-09-2018

Union or the Council of Europe. In fact, this period is characterised by a certain balance at an institutional level: the European Movement had two operational headquarters (London and Paris), the Council of Europe was founded in London but seated in Strasbourg, UNESCO settled in Paris but its first Director-General (Julian Huxley) was British, the education conferences promoted by the Western Union were first held in Great Britain and afterwards in France…

Regarding cultural supremacy, France laid out different soft power resources on the table, such as its prestigious public secular educational system, its renowned universities, the wide presence of the French language at school, its innovations in the field of language teaching through the Direct Method… Similarly, the country deployed its intelligent power resources: it combined the prestige of its national culture with hard power elements associated with its *grandeur*, by using the intergovernmental organisations we have analysed to spread its relational power with the aim of turning the French culture into a model for the European. All this enabled the country to have a relevant presence in the agenda of the European Cultural Conference (1949) and the UNESCO seminar held in Nawara Eliya, but it was not enough to impose its will, for instance, with regard to introducing a bilingual system (English-French) at schools or to giving its national culture the status of leader of the European culture.

Great Britain, for its part, employed different soft power resources such as the global expansion of its language –linked to North America's supremacy–, the prestige of its universities and its educational system –decentralised and considerably autonomous, for instance, concerning the choice of textbooks– and its innovations in the educational field with regard to the Audio-lingual Method (also known as the Army Method) –shared, in this case, with the USA–, applied to language teaching. The British intelligent power becomes allies with the American, which invests great sums through the *Network State-Private* to spread the Anglo-Saxon culture. All these factors contribute to a change in the role played by English at school (since it will become a predominant foreign language), as well as to avoid the approval of a European History textbook blessed by the Council of Europe in opposition to the traditional autonomy on the part of institutions when selecting teaching materials.

The Federal Republic of Germany, a newborn country in 1949, was also able to use soft power resources in the dispute over cultural hegemony, such as the significant influence of its language in Europe,

Received: 09-11-2017 Accepted: 14-09-2018

the prestige of its universities and its considerable cultural heritage prior to the emergence of Nazism. Moreover, the intelligent power linked to its convenient geographical location in the centre of Europe was also useful with regard to the debate on whether feudalism should be considered one of the principles which inspired the idea of Europe. Similarly, its support of Europeanism, clearly reflected on the contents of its History textbooks, enables to control the aspirations of those who desire a bilingual educational system as well as to exert an especially active relational power in the sphere of cultural diplomacy.

With regard to small countries, they will also make use of their resources to participate in this cultural, educational conflict. Scandinavian countries, for example, will use both their soft power (cooperation in the educational field) and relational power (creation of the Nordic Council in 1952). As for Greece, it takes advantage of the great prestige of its cultural heritage to act as a representative of the European periphery (in opposition to the Nordic-Carolingian model) concerning issues such as how to materialise the idea of Europe and how to spread it at school.

Finally, not only the Western Union but also the European Movement and the Council of Europe will find out that school and education are a fertile soil to legitimise their aspirations of unity, which can indeed serve as a clear example of their relational power.

References

Bayet, J. (1949), *Rapport à la Conférence Culturelle de l´Europe Unie sur l'*éducation par M. Jean Bayet. Archives Historiques de l´Union Européenne (AHUE). Folder ME-540.

Bossuat, G. (1994), *Fondateurs de l´Europe*, Paris: Belin.

Bossuat, G. (2005), *Faire l´Europe sans défaire la France*, Bruxelles: Peter Lang.

Council of Europe (1952a), *Les actitivités culturelles du Conseil de l´Europe. Mémorandum présenté au Comité des Experts culturels par le Sous-Comité Consultatif des Cinq, Strasbourg, 24-28 avril 1952*. Retrieved from https://rm.coe.int/16804e17e7#search=%22Comite%20

des%20experts%20culturels%22, amendments available at http://coe.archivalware.co.uk/awweb/pdfopener?smd=1&md=1&did=938985

Council of Europe (1952b), *Assemblée Consultative. Commission des questions culturelles et scientifiques. Procès-verbal de la séance tenue à Strasbourg le jeudi 29 mai 1952*. Retrieved from http://coe.archivalware.co.uk/awweb/pdfopener?smd=1&md=1&did=676729

Council of Europe (1953a), *Programme du symposium sur la révision des manuels d´histoire, Calw (Forêt-Noire) du 4 au 12 août 1953*. Retrieved from http://coe.archivalware.co.uk/awweb/pdfopener?smd=1&md=1&did=960396

Council of Europe (1953b), *Compte-rendu du symposium sur la révision des manuels de histoire qui s´est tenu à Calw (Fôret-Noire) du 4 au 12 de août 1953*. Retrieved from http://coe.archivalware.co.uk/awweb/pdfopener?smd=1&md=1&did=496234

Davies, N. (1996), *Europe. A history*, Oxford: Oxford University Press.

Delgado Gómez-Escalonilla, L. (2015), "Los dilemas de un propagandista reticente. El caso de Estados Unidos" in Gavari Starkie, E. & Rodríguez Jiménez, F.J. (coords.), *Estrategias de diplomacia cultural en un mundo interpolar*, pp. 47-94. Madrid: Editorial Universitaria Ramón Areces.

Evans, I. (1948), *Carta de 19 de marzo de 1948 de Ifor Evans a Kenneth Lindsay*. AHUE. Instituto Europeo de Florencia. Folder ME-390.

Flick, U. (2004), *Introducción a la investigación cualitativa*, Madrid: Morata.

García Pedraza, R. (2015), "Desafíos a la igualdad en el siglo XXI: dificultades y retos del modelo de escuela comprensiva", *Revista Española de Educación Comparada*, 26, 135-152.

Gavari Starkie, E. (2015), "La política educativa: ejemplo del *soft-power* europeo" in Gavari Starkie, E. & Rodríguez Jiménez, F.J. (coords.), *Estrategias de diplomacia cultural en un mundo interpolar*, pp. 31-46. Madrid: Editorial Universitaria Ramón Areces.

Gómez Carrasco, C.J. & Chapman, A. (2017), "Enfoques historiográficos y representaciones sociales en los libros de texto. Un estudio comparativo, España-Francia-Inglaterra", *Historia y memoria de la educación*, 6, 319-361.

Haigh, A. (1974), *Cultural diplomacy in Europe*, Strasbourg: Council of Europe.

Judt, T. (2007), *Pasado Imperfecto. Los intelectuales franceses (1944-1956)*, Madrid: Taurus.

López Facal (2010), "Nacionalismos y europeísmos en los libros de texto: identificación e identidad nacional", *Clio & Asociados*, 14, 9-33.

Morales Gil, F.J., (2007), "La difusión de la metodología activa en la enseñanza de los idiomas en España: el caso del francés", *XXI. Revista de Educación*, 9, 237-251.

Mouvement Européen, (1949), *Conférence Européenne de la Culture. Résolutions et déclaration finale*. Lausanne: Mouvement Européen.

Nye, J. (1990), *Bound to lead: the changing nature of American power*, New York: Basic Books.

Nye, J. (2004), *Soft Power: the means to success in world politics*, New York: Public Affairs.

Nye, J. (2011), *The Future of Power*, New York: Public Affairs.

Puren, Ch. (2012) [1988], *Histoire des méthodologies de l´enseignement des langues vivantes*, Paris, Nathan-CLE International. Retrieved from https://www.aplv-languesmodernes.org/docrestreint.../puren_histoire_methodologies.pdf

Rougemont, D. (1949), *Présentation du Rapport Général par Denis de Rougemont (discours prononcé le 8 décembre 1949 à la première séance plénière)*, AHUE. Folder ME-538.

Sartre, J-P. (1949), *Défense de la culture française par la culture européenne*, Paris : Centre d´Études de Politique Étrangère. AHUE. European University Institute. Folder ME-528.

Scott-Smith, G. & Krabbendam, H. (eds.) (2004), *The Cultural Cold War in Western Europe, 1945-1960 (Studies in Intelligence)*, London: Routledge.

Saunders, F. S. (1999), *Who Paid the Piper? The CIA and the Cultural Cold War*, London: Granta.

UNESCO (1950), *Better History Textbooks*, Paris: UNESCO.

UNESCO (1951a), *International educational seminar on the teaching of history as a means of developing international understanding. Final report of the study group II*. Sèvres: UNESCO. Retrieved from http://unesdoc.UNESCO.org/images/0015/001556/155685eb.pdf

UNESCO (1951b), *International educational seminar on the teaching of history as a means of developing international understanding. Final report of the study group III*. Sèvres: UNESCO. Retrieved from http://unesdoc.UNESCO.org/images/0015/001556/155684eb.pdf

UNESCO, (1954), *An account of the international seminar on the contribution of the teaching of modern languages towards education for living in a World community*, Paris: UNESCO. Retrieved from http://unesdoc.UNESCO.org/images/0012/001261/126139eb.pdf

Vigander, H. (1950), *Mutual revision of History textbooks in the Nordic Countries*, Paris: UNESCO.

Contact address: Luís Domínguez Castro. Universidad de Vigo. Facultad de Filología y Traducción. Departamento de Historia, Arte y Geografía. Campus Lagoas-Marcosende, s/n 36210. Vigo. E-mail: dominguez@uvigo.es

Received: 09-11-2017 Accepted: 14-09-2018

The assessment of narrative competence during compulsory schooling[1]

La evaluación de la competencia narrativa en la educación básica

DOI: 10.4438/1988-592X-RE-2019-383-402

María-Jesús Fernández
Manuel Montanero
Manuel Lucero
Universidad de Extremadura

Summary

Story writing is a fundamental activity toward the development of writing skills in the first years of school education. In a classroom context however, its assessment is often focused on superficial aspects of grammar and spelling. It is widely known that there is a lack of instruments with demonstrated validity and reliability to enable the assessment of semantics and rhetoric in this context. Therefore, the objective of this research was to build and validate a story assessment rubric (RER) to assess the writing tasks of school children of 8-14 years of age. We analyzed 292 texts written by students at those levels of education. The data provided support the validity of the story assessment rubric (RER). This suggests that the dimensional structure, as well as the operationalization of criteria and levels of achievement, are relevant and discriminative enough to be implemented at these levels of the education system. The reliability of the instrument also yielded positive results, in terms of internal consistency and objectivity of implementation. However, the results of the pilot study show some small inconsistencies in a few levels of achievements, such as in the highest

(1) Acknowledgements: To the Consejería de Economía, Competitividad e Innovación of the Regional Government of Extremadura, and to the European Fund of Regional Development for financial support received (Reference: GRU15153).

Received: 27-04-2018 Accepted: 05-10-2018

level of the "plot" criterion. This is due to the insufficient number of secondary education students capable of writing complex and articulate plots with truly creative elements. Finally, these results are discussed, as well as their implications toward the assessment and improvement of narrative skills.

Key words: Writing; Narrative Skills; Stories; Rubric; Elementary Education; Secondary Education.

Resumen

La redacción de relatos es una actividad fundamental para el desarrollo de la competencia escritora en los primeros años de escolaridad. Sin embargo, su evaluación en el contexto del aula se centra frecuentemente en aspectos superficiales, de carácter gramatical y ortográfico. Es notoria la escasez de instrumentos funcionales, con suficientes garantías de validez y fiabilidad, que faciliten la evaluación de aspectos semánticos y retóricos en la práctica del aula. En este sentido, el objetivo de esta investigación fue construir y validar una rúbrica para evaluar relatos (RER) redactados por estudiantes de entre 8 y 14 años. Para ello, se analizaron un total de 292 textos producidos por estudiantes escolarizados en estos niveles educativos. Los datos aportados apoyan la validez de la RER, lo que sugiere que tanto la estructura dimensional como la operativización de los criterios y niveles de logro son relevantes y suficientemente discriminativos para su aplicación en estos niveles educativos. También se obtuvieron datos positivos de la fiabilidad del instrumento, desde el punto de vista de su consistencia interna y de la objetividad de su aplicación por diversos evaluadores. Los resultados del estudio-piloto muestran, sin embargo, pequeñas inconsistencias en algunos niveles de logro, como el nivel máximo del criterio "trama". Esto se debe a la escasez de estudiantes de Secundaria que consiguen redactar tramas complejas, bien articuladas y con elementos realmente creativos. Finalmente, se discuten estos resultados y sus implicaciones para la evaluación y mejora de la competencia narrativa.

Palabras-clave: Escritura; Competencia narrativa; Relatos; Rúbrica; Educación Primaria; Educación Secundaria.

Introduction

Story writing is a fundamental educational activity toward the acquisition of writing skills during the first years of schooling. From the age of 5, children are already able to build simple temporal sequences verbally,

Received: 27-04-2018 Accepted: 05-10-2018

contextualizing them in space-time frames, and then progressively adding causal and motivational relations. From the age of 7, most children are gradually able to put those stories in writing. The characters are introduced by their names or from the narrator's perspective, but are scarcely described (Kernan, 1977); the actions are articulated in episodes and the stories become progressively more coherent, to the extent that the listener can predict a logical ending during the narration (Stadler & Ward, 2005). The sequences are no longer only showing causal relations between certain actions and the characters' reactions, but they are rather clearly articulated around a problem, a solution, and a mediating element between them (Iandolo, Esposito & Venuti, 2012). Between the ages of 8 and 9, there is a significant improvement in cohesion and complexity of the story's structure, incorporating more problematic and mediation elements that reinforce the problem itself (Stein & Glen, 1979). Around the ages of 10 or 11, the narrative essays can be considered relatively mature (Artiles & Jiménez, 2007; Yan *et al.*, 2012), with a hierarchical structure of episodes (primary and secondary) articulated around an inciting incident or a central topic (Botvin & Sutton-Smith, 1977; Ochoa-Angrino *et al.*, 2008).

This progression is determined by the cognitive level, language proficiency and the individual's emotions (Ryan, 2014). For example, the compositions of children with some deficiency in motor skills required for the writing skills, or those who make many spelling mistakes, demonstrate a worse organization of the text or textualization (Berninger *et al.*, 1992; Cremin & Myhill, 2012). When those processes become more automatic, the compositions also improve in other aspects (Graham, Harris & Fink-Chorzempa, 2002).

For decades, we have had a relatively broad knowledge of the evolution of the children's narrative skills. However, when assessing stories, teachers usually neglect the semantic content and rhetorical structure of the narration (Fernández, Lucero & Montanero, 2016). Rather, the frequency of grammar and spelling corrections recorded in students' manuscripts has proven to be the best predictor of the overall grade that teachers give to the texts (Lucero, Fernández & Montanero, 2018). This could be the reason why elementary school children substantially improve those last aspects in their drafts after receiving their teachers' assessment, although not the content and rhetorical structure (Montanero, Fernández & Lucero, 2014).

Received: 27-04-2018 Accepted: 05-10-2018

Although the genre of narrative texts is the most frequent in reading and writing assignments in elementary education, it is striking that very few of the most popular standardized tests for its assessment take rhetorical skills into consideration (see Galve, 2005; García & Marbán 2001; Toro & Cervera, 1984; Toro, Cervera & Urío, 2002). In Spain, the assessment test of writing processes (PROESC) (Cuetos, Ramos & Ruano, 2002) is practically the only standardized test that allows for the assessment of these skills. Its use in the classroom, however, is marginal, probably due to its cost.

It is reasonable to think that if teachers had practical assessment tools in the classroom, the quality of feedback offered to students would improve. Thus, the use simpler instruments, based on scales or rubrics, has been proposed (Sundeen, 2014). A rubric is a particular type of ordinal measurement, in a descriptive format. Its criteria are operationalized in various levels of achievements. This characteristic allows the students to know their situation regarding the objectives they must reach within a specific task.

For example, Benítez & García (2010) proposed an assessment scale of six values per criteria (from unsatisfactory to excellent), allowing the assessment of rhetorical elements in the stories, such as the identification of characters or construction of the plot.

Yan *et al.* (2012) devised a writing assessment rubric, composed of seven criteria divided in two dimensions (content and organization), for students of 6-9 years of age. The instrument valued relevance, range, the use of connectors, coherence, text structure, and intelligibility.

Dunsmuir *et al.* (2015) proposed another rubric allowing for the assessment of written texts in children between 7-11 years of age; this rubric had seven criteria and four levels. The aspects taken into account in this case were the calligraphy, spelling, punctuation marks, structure of the text, vocabulary, organization and originality.

It has been proven that these last instruments enable teachers to explain to the students which aspects will be considered in the assessment, and provide more information than the usual correction of compositions (Ferris & Hedgcock, 2014; Li & Lindsey, 2015). The necessity to select one of those levels for each criteria of the rubric triggers a richer interaction in situations of co-assessment, which in turn has a significant influence in the quality of narrative compositions (Montanero *et al.*, 2014; Ochoa-

Received: 27-04-2018 Accepted: 05-10-2018

Angrino *et al.*, 2008), the self-regulation of learning and the feeling of self-efficacy (Panadero & Jonsson, 2013).

However, other studies have highlighted certain weaknesses related on the one hand, to its validity, and on the other, to its limitations in the formative assessment of competences (Spandel, 2006). About the former, Jonsson & Svingby (2007) did a review of the validity and reliability of various rubrics in different studies. Only a third of these gave some information as to the validity of the instruments. Most validity indicators are based on expert opinions, confirmatory factor analysis (Baker *et al.*, 1995), or on correlations (greater than 0,40) with other measurements taken as criteria (Roblyer & Wiencke, 2003). As regards reliability, the rubrics reviewed by Jonsson & Svingby yielded values between 0,50 and 0,92 in the Cronbach's Alpha test; a 55%-75% degree of agreement between judges; and a Cohen's kappa coefficient between 0,4 and 0,75. Only one of the reviewed studies (Gearhart *et al.*, 1995) was focused on analyzing the validity of a story assessment rubric.

Regarding the second question, the use of the rubric in situations of formative assessment of writing skills, some authors pointed out the risk that the instrument might hinder the students' writing performance (Fang & Wang, 2011). Many authors argue that although they enable teachers to do a quick assessment, they do not really contribute to improve students' writing skills. This is due to the fact that the achievement levels involve generic (not personalized) suggestions, and they are often formulated in a language that is difficult for students to understand (Janssen *et al.,* 2015; Kohn, 2006; Panadero, Alonso-Tapia & Huertas, 2012; Schenck & Daly, 2012). Finally, they also pointed out the lack of training of some teachers who are using it in a classroom context (East, 2009; Knoch, Read & von Randow, 2007).

Ultimately, it seems necessary to shed more light on the validity and usefulness of this type of classroom instruments, particularly as a support for the assessment of writing skills in elementary and secondary education. Within this framework, the objective of this study was to devise a narrative writing assessment tool that would be useful and practical for teachers in the classroom. We built and validated an "analytical" rubric enabling a quick and simple formative assessment of narrative texts written by students of 8-14 years of age. To this effect, we will provide more evidence of validity and reliability than what has been published so far in relation to the abovementioned rubrics. We will also analyze the

Received: 27-04-2018 Accepted: 05-10-2018

individual differences observed when using the rubric with a sample of students of different ages. In doing so, we will be able to verify if the rubric allows us to distinguish between different levels of narrative skills in students' compositions in relation to their level of education.

Methods

Sample

A total of 292 students participated in the study, ranging from the third year of elementary education to the second year of compulsory secondary education (8-14 years old). The sample was obtained from seven schools of different cities in Extremadura, selected for convenience in areas of low or average to low socio-economic level. Immigrant students with an insufficient level of Spanish did not participate, nor students with special educational needs. Table 1 shows the final distribution of the sample by age, gender, and level of schooling.

TABLE 1. Sample distribution (N) according to the level of education (EP: elementary school; ESO: compulsory secondary education), age (in years) and gender (M: male; F: female)

Level	School year	Age	Gender	N		
I	3º EP	8-9	M	14	40	92
			F	26		
	4º EP	9-10	M	17	52	
			F	35		
II	5º EP	10-11	M	18	49	102
			F	31		
	6º EP	11-12	M	25	53	
			F	28		
III	1º ESO	12-13	M	11	25	98
			F	14		
	2º ESO	13-14	M	30	73	
			F	49		

Prepared by the authors

Received: 27-04-2018 Accepted: 05-10-2018

Instruments

Writing sheet

Each student was provided with a writing sheet, on which they had to write a short story. The topic was indicated at the top ("a child who becomes the hero of his village"), as well at the required length (approximately one page).

Story assessment rubric

We adapted an instrument which had already been used in previous studies (Montanero, Lucero & Fernández, 2014), comprising seven assessment criteria, distributed in two dimensions. On the one side, those criteria evaluated aspects related to content and narrative organization: Frame, Topic, Plot, and Creativity. On the other side, grammar, lexical and orthographic aspects were evaluated: Sentences, Vocabulary, and Spelling (Table 2). Each of those criteria had four levels of accomplishment or execution, which could be rated as such: 0 points (level 1), 0,5 points (level 2), 1 point (level 3) and 1,5 points (level 4).

PROESC

The test of writing processes (PROESC) (Cuetos, Ramos & Ruano, 2002) is a standardized test, commercialized by TEA, which is widely used in the narrative writing assessment in compulsory education. It is composed of ten assessment criteria grouped into two main dimensions: *content* and *coherence*. The first dimension focuses on the following criteria: Where and When, Characters, Events and Consequences, Coherent Ending, and Creativity. The second dimension is assessed according to the following criteria: Logical Continuity, Sense of Unity, Literary Figures, Complex Sentences, and Vocabulary. To achieve each of those criteria, many requisites are defined.

According to the manual, the instrument has an internal consistency of 0,82 (Cronbach's Alpha coefficient). As evidence of criterion validity,

the application of the test to a sample of 839 texts yielded a correlation of 0,46 with the grades given by a group of teachers. Regarding the structural validity, the ten criteria which constitute the test were submitted to a components analysis with varimax rotation. The result was the division of the test in three components which explained 58% of the total variance.

Procedure

Design of the story assessment rubric

The object of this study, the rubric, consists of a descriptive-ordinal assessment scale (with four levels per criteria, which were also given a numerical score), with the purpose of assessing stories written in L1 by students of 9-14 years of age. Although the instrument was initially intended for all compulsory education students, a preliminary study (Fernández, 2017) concluded that it was not valid nor discriminative enough for the early elementary school students (6-8 years old), or for the last years of secondary school (15-16 years).

The assessment criteria of the instrument were based on theorical variables chosen after reviewing previous studies on the assessment of narrative skills (Benítez & García; 2010; Cuetos *et al.*, 2002; Dunsmuir *et al.*, 2015; Montanero *et al.*, 2014; Sundeen, 2014; Yan *et al.*, 2012; Thorndyke, 1977).

From this bibliography review, seven quality criteria for a narrative composition were identified. The first four are focused on the assessment of the story's organization and content; the others focus on the grammar, lexical and orthographic aspects.

- *Frame*. Corresponds to the presentation of a space-time frame, which can be more or less defined and identifiable, or fictitious, and in which the characters are usually presented. The most important aspect is the richness of character descriptions (from a physical or psychological viewpoint).
- *Topic or inciting incident*. This is the inciting incident which triggers all subsequent actions. A good story refers to the emotions

Received: 27-04-2018 Accepted: 05-10-2018

which make the characters react, as well as the intentions guiding their actions.

- *Creativity and interest*. It is related to originality, creativity and the length of the story. It is highly valued when the story is no mere reproduction of a known tale, if it has an adequate length, and whether the author uses certain stylistic resources such as metaphors, moral of the story, sense of humor, etc.
- *Sentences*. It corresponds to sentence structure and the use of punctuation in the text. An adequate story is characterized by the intelligibility of its sentences, by having an appropriate structure, and by a suitable use of punctuation marks.
- *Vocabulary*. Refers to the variety and correctness of the words the authors include in their story, without ambiguity or excessive repetitions.
- *Spelling*. Refers to the correct use of the grammar rules (in their mother tongue).

Subsequently, a draft of the rubric was drawn up, in which we defined the four levels of achievement for each of the previously mentioned criteria expected from writers of 10-12 years of age.

Finally, the story assessment rubric was applied in various cycles to about 30 compositions of students of this level of education. At the end of each cycle, the researchers re-defined the formulation of some levels of achievement, to improve clarity and objectivity, as well as its capacity of discrimination (mostly to avoid the risk of floor effect in some criteria). Table 2 shows the final version of the instrument.

Story writing and assessment

The validation of the story assessment rubric (RER) was based on its application to a sample of narrative texts of 292 students of 8-14 years of age.

The text writing process happened during one hour-long session. In a first phase of 20 minutes, the concept of narration was briefly introduced, as well as the characteristics and typical structure of a story. In the second phase, each student had 30 minutes to write an invented story, with the

topic *a child who becomes the hero of his village*, of a length of one or two pages.

The narrative texts produced by the children were analyzed from three perspectives. On the one hand, we applied independently the story assessment rubric (RER) and the PROESC. In both cases, each composition received the score corresponding to a criteria or level when the text met all the requirements. If any of those were missing, the next lower score was obtained.

On the other hand, four teachers with more than five years of professional experience were selected (in other schools than those where the study was carried out). These teachers graded the compositions using a scale between 0 and 10, without knowledge of the score achieved previously by the story assessment rubric and the PROESC.

Results

Next, we will present the results of the validity and reliability analysis for the rubric we devised to assess narrative texts.

Content validity

The rubric was submitted to a group of experts to evaluate its content validity. Two scholars were selected, specialists in writing, with an experienced professional in elementary and secondary education. They were sent a validation sheet in which they were asked to assess, on the one hand, the suitability of the selected criteria and their importance regarding the quality of a story; on the other hand, the clarity of formulation of the achievement levels.

The average obtained in the evaluation of importance (on a scale from 1 to 5) was 5 in all criteria, except in Creativity (4,6), Sentences (4,3), and Vocabulary (4,6). In terms of clarity of formulation, only the highest score was obtained, except for the Topic (4,6). Furthermore, the experts made some minor observations, which were taken into account for the revision of the final draft of the story assessment rubric (Table 2).

Received: 27-04-2018 Accepted: 05-10-2018

TABLE II. Story Assessment Rubric (RER)

Criteria	Levels of Achievement	
1. Frame (Introduction)	1. There is no introduction to the story	
	2. Only a few characters are named, without mentioning anything else about them.	
	3. The story tells where and when the action happened and who are the protagonists, but they are not described any further.	
	4. The story tells where and when the action happened and who are the protagonists; their physical aspect and personality is also described.	
2. Topic or Inciting Incident	1. What happened at the beginning is not explained.	
	2. What happened at the beginning is partially explained, but it is not comprehensible.	
	3. What happened at the beginning is partially explained, as well as how the characters felt.	
	4. What happened at the beginning is clearly explained, as well as how the characters felt and what they planned to do.	
3. Plot (episodes and resolution)	1. Nothing happens in the story or what happens is not comprehensible.	
	2. The story is a bit confusing or some facts do not make much sense.	
	3. The story is simple. What happened next and at the end is explained with enough clarity.	
	4. Although the story is long, what happened is very clear (in well-organized segments), as well as what happened to the characters at the end.	
4. Creativity and Interest	1. The story is copied or cannot be understood.	
	2. The story is original and can be understood, but it is too short or very boring.	
	3. The story is original and entertaining.	
	4. The story is original and entertaining. It is nice, shows a sense of humor, or teaches us something.	
5. Sentences	1. Most sentences cannot be understood because they are badly built.	
	2. Most sentences can be understood, but some are badly built.	
	3. The sentences are easily understood, but there are too few punctuation marks (or there are only commas).	
	4. The sentences are easily understood, most of the punctuation marks are in the right places, even in the dialogues.	
6. Vocabulary	1. The vocabulary is very poor and with many mistakes (some sentences are not correctly representing what they are trying to say).	
	2. The words are correct, but there are a lot of repetitions.	
	3. The vocabulary used is correct and varied (not repeated too often).	
	4. The vocabulary used is rich, with some little-known words.	
7. Spelling	1. There are a lot of spelling mistakes (at least one in each sentence).	
	2. There are quite a lot of spelling mistakes (at least one in every other sentence).	
	3. There are few spelling mistakes.	
	4. There are no spelling mistakes.	

Adapted from Montanero, Lucero & Fernández (2014)

Construct validity: analysis of the individual differences according the level of education

The final version of the story assessment rubric was applied to a sample of 292 students of elementary and secondary education. Table 3 reflects the average results for each criterion in the three levels of education that were compared. Between the first and second level, there were no significant differences in terms of the frame (t=2,48; p<0,05), the topic (t=3,07; p<0,01), the sentences (t=2,69; p<0,01), the spelling (t=4,48; p<0,01); as well as in terms of the subtotal of the Organization and Content dimension (t = 3,18; p<0,01), the subtotal of the grammar, lexical and orthographic aspects dimension (t = 4,25; p<0,01), and for the total of the story assessment rubric (t=4,42; p<0,01). Significant differences were detected between the second and third level, in terms of creativity (t=3,52; p<0,01) and spelling (t=2,58; p<0,05); as well as in the subtotal of Organization and Content (t=3,18; p<0,01), the subtotal of Grammar Aspects (t=2,58; p<0,05), and the total for the rubric (t=3,50; p<0,01).

TABLE 3. Differences in results of the story assessment rubric (RER), according to the level of schooling (LI: third and fourth year of elementary education; LII: fifth and sixth year of elementary education; LIII: first and second year of compulsory secondary education).

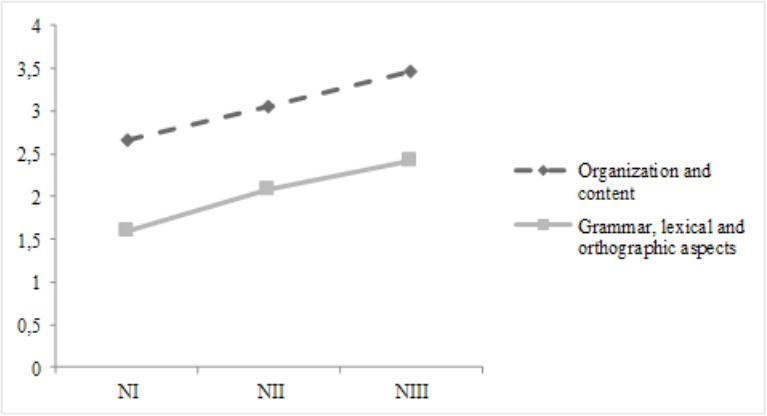
RER	**Mean (M) y standard deviation (SD)**						**Difference of means (DM)**		
Criterion	*Level I*		*Level II*		*Level III*		*LI-II*	*LI-III*	*LII-III*
	M	*SD*	*M*	*SD*	*M*	*SD*	*DM*	*DM*	*DM*
Frame	0,79	0,4	0,92	0,35	0,99	0,38	-0,13*	-0,2**	-0,07
Topic	0,63	0,31	0,77	0,34	0,87	0,37	-0,14*	-0,24**	-0,1
Plot	0,77	0,32	0,83	0,37	0,89	0,28	-0,06	-0,12*	-0,06
Creativity	0,47	0,32	0,54	0,38	0,7	0,28	-0,07	-0,23**	-0,16*
Subtotal Organization y Content	2,65	0,87	3,06	0,95	3,46	0,8	-0,41**	-0,81**	-0,4**
Sentences	0,67	0,33	0,81	0,39	0,91	0,41	-0,14*	-0,24**	-0,1
Vocabulary	0,63	0,35	0,71	0,35	0,8	0,45	-0,08	-0,17**	-0,09
Grammar	0,28	0,45	0,55	0,40	0,7	0,4	-0,27**	-0,42**	-0,15*
Subtotal grammar, lexical, and orthographic aspects	1,59	0,75	2,08	0,85	2,41	0,95	-0,49**	-0,82**	-0,33*
Total	4,23	1,35	5,14	1,50	5,87	1,43	-0,91**	-1,64**	-0,73**

Prepared by the authors. (*) p<0,05; (**) p<0,01

Received: 27-04-2018 Accepted: 05-10-2018

The validity indicator of an instrument is its capacity to discriminate between the different levels of development of the competence it is supposed to assess. The next graphic shows a progressive and linear increase of the rubric's average as the students' level of schooling progresses.

GRAPH 1. Evolution of global criteria in terms of level of schooling (LI: third and fourth year of elementary education; LII: fifth and sixth year of elementary education; LIII: first and second year of compulsory secondary education).



In a more specific way, Table 4 shows the distribution of subjects for each level of achievement, according to the criteria of the story assessment rubric (RER).

As predicted, the *content* of stories becomes more complex as students advance through the different levels of the education system. For example, a third grade student does not usually refer to the *space-time* frame of the text. In the last years of elementary education, on the other hand, the name of the place is often mentioned, and they use specific genre formulas to refer to the temporal context (e.g. "Once upon a time"). In elementary education, in general, the narration almost always revolves around a single character who is scarcely described, or if there is a description, it is only to point out some physical characteristic. Sometimes, there is a mention of a character which was unknown until then, or whose role in the plot is not entirely clear. In secondary education, in contrast, a context is given to the story and the characters

are described in more details, including psychological and psychosocial traits ("Luis wasn't a good student, but he was quiet and he was a team player"). Additionally, *secondary characters* start appearing and participate in relevant events.

TABLE 4. Percentage of students in each level of achievement (LoA) of the story assessment rubric (RER) according to the level of schooling (LI: third and fourth year of elementary education; LII: fifth and sixth year of elementary education; LIII: first and second year of compulsory secondary education).

Criterion	LoA	NI	NII	NIII
1. Frame	1	1,1	1	0
	2	58,7	30,4	28,6
	3	21,7	52	43,9
	4	18,5	16,7	27,6
2. Topic or inciting incident	1	5,4	1	0
	2	68,5	53,9	42,9
	3	21,7	35,3	39,8
	4	4,3	9,8	17,3
3. Plot	1	2,2	4,9	0
	2	47,8	35,3	29,6
	3	44,6	48	63,3
	4	5,4	11,8	7,1
4. Creativity	1	22,8	21,6	3,1
	2	62	52	53,1
	3	14,1	23,5	43,9
	4	1,1	2,9	0
5. Sentences	1	7,6	6,9	2
	2	52,2	35,3	38,8
	3	38	46,1	34,7
	4	2,2	11,8	24,5
6. Vocabulary	1	12	5,9	10,2
	2	52,2	52	36,7
	3	33,7	36,3	35,7
	4	2,2	5,9	17,3
7. Spelling	1	68,4	25,5	15,3
	2	8,7	39,2	33,7
	3	20,7	34,3	46,9
	4	2,2	1	4,1

Prepared by the authors.

Received: 27-04-2018 Accepted: 05-10-2018

The story assessment rubric also shows how the *plot* becomes more complex throughout the different levels of education. In many third grade texts, almost nothing happens (they are simple descriptions without problematic events to trigger subsequent actions); while in others, the plot is confusing, with disruptions in the common thread, and the ending is often not coherent with the beginning of the story. At this level, many children do not write an original story, but rather adapt elements of well-known tales. A significant improvement in coherence can be observed from the fifth grade onward. The characters' feelings are now explicitly mentioned, as well as what they plan to do once the initial problem has been delimited. In many texts though, the actions are still juxtaposed, and there can be a lack of causal or motivational logic. The stories are longer and more creative, and often incorporate dialogues and morals. In the first years of secondary school, the plot is extended, incorporating more episodes, and there are sometimes multi-problem topics which generate cycles of denouement.

Surprisingly, the percentage of those secondary school students reaching the highest level of achievement is slightly lower than the students in their last years of elementary education. The texts from secondary school students are longer, but not much more creative than in the previous level of education, and moreover they usually lack dialogues and other enriching elements.

Regarding the *formal* and *grammar* aspects, the results of the story assessment rubric are in line with the expected progress according to the students' level of education. In the texts of third and fourth grade students, the sentences are mostly simple, or the compound sentences are badly coordinated (with an excessive use of the anaphora, as well as the conjunction "and"). 70% of third and fourth grade students made more than ten mistakes per text (not including the Spanish *tilde*); while most fifth and sixth grade students made more than five. Substantive and adjective subordinate clauses practically do not exist until secondary school, and the punctuation marks are used poorly (although there is a systematic abuse of commas where there should be a period or a semi-colon). The vocabulary and the anaphoric references are still limited and repetitive for most students at this level of education. Almost all texts in secondary education still contain spelling mistakes, although for most of them, less than five.

Received: 27-04-2018 Accepted: 05-10-2018

Normality

The following table shows the scores obtained by the compositions for the different levels of education, and for each of the dimensions. They are sufficiently adapted to the normal distribution models, taking into account in each case the goodness of fit, asymmetry, and kurtosis.

TABLE 5. Score distribution for the story assessment rubric (RER) per dimension for each level of schooling (LI: third and fourth year of elementary education; LII: fifth and sixth year of elementary education; LIII: first and second year of compulsory secondary education).

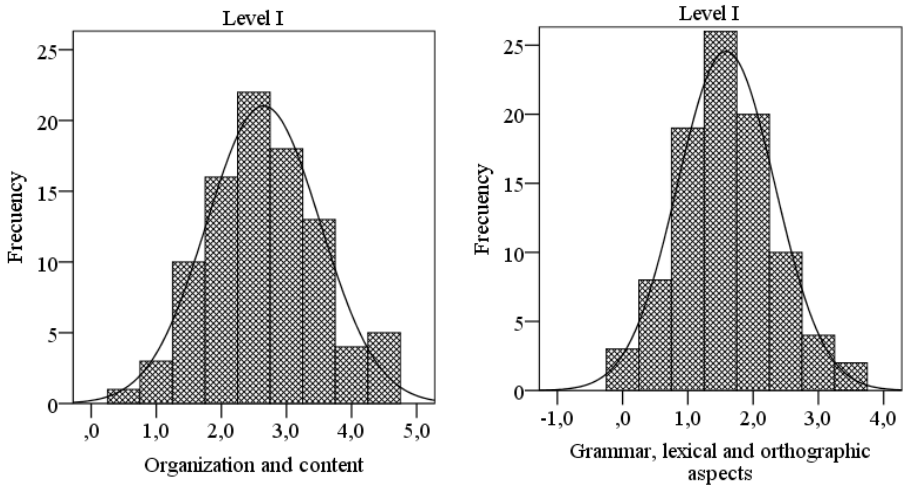
Dimensions	Level I	Level II	Level III
Organization and Content	Mean: 2,65 SD: 0,87 Asymmetry: 0,13 Kurtosis: -0,20 Z of K-S: 1,27	Mean: 3,06 SD: 0,95 Asymmetry: 0,14 Kurtosis: -0,59 Z of K-S: 1,48	Mean: 3,46 SD: 0,80 Asymmetry: 0,06 Kurtosis: -0,67 Z of K-S: 1,43
Grammar, lexical y orthographic aspects	Media: 1,59 DT: 0,75 Asymmetry: 0,22 Kurtosis: 0,01 Z of K-S: 1,49	Media: 2,08 DT: 0,85 Asymmetry: 0,10 Kurtosis: 0,03 Z of K-S: 1,36	Media: 2,41 DT: 0,95 Asymmetry: 0,05 Kurtosis: -0,51 Z of K-S: 1,24

Prepared by the authors

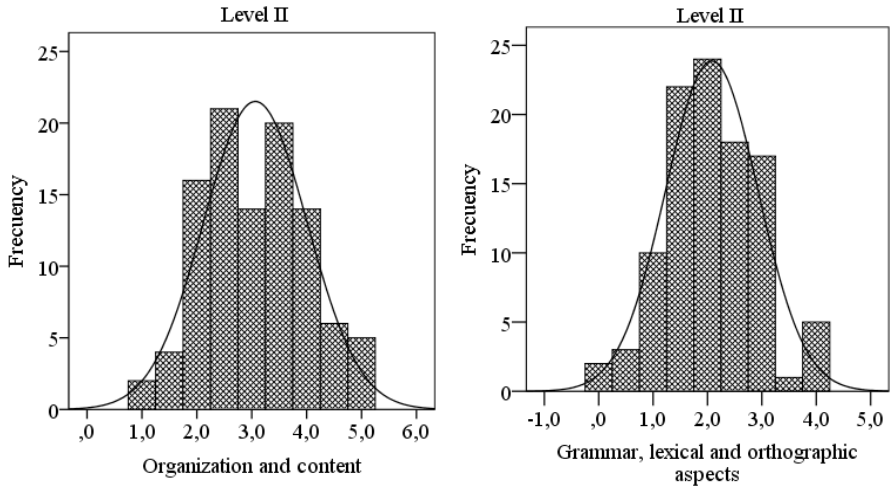
Graphs 2, 3, and 4 show a better adjustment in levels of education I and II. In the Organization and Content dimension, the curve is slightly platykurtic in all levels, but especially in level III.

Received: 27-04-2018 Accepted: 05-10-2018

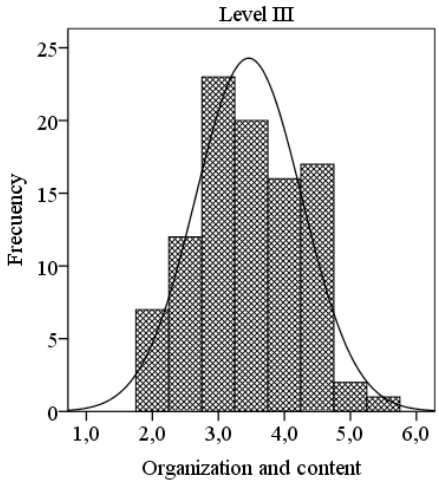
GRAPH 2. Score distribution in Level I (third and fourth grade)

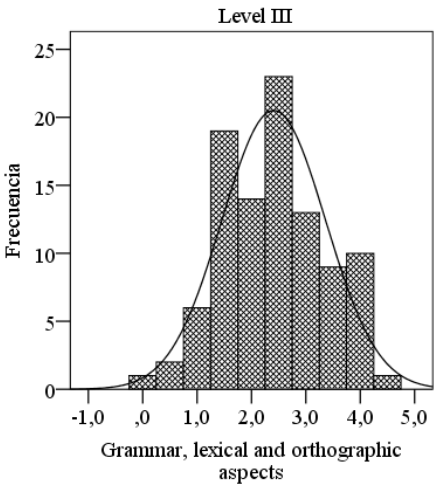


GRAPH 3. Score distribution in Level 2 (fifth and sixth grade)



GRAPH 4. Score distribution in Level III (first and second year of secondary education)





Structural validity

A confirmatory factor analysis of the main components of the rubric was carried out in order to evaluate the dimensions and criteria of the story assessment rubric. Two tests were previously performed, which confirmed that the data set could be factorized: the Kaiser-Meyer-Olkin measure of sampling adequacy (0,75) and Bartlett's test for sphericity (Chi-squared=161,45; p<0,01).

Table 6 shows the factor matrix of rotated components, according to the varimax procedure. Both extracted components can be observed, as well as the saturation of variables for each of them.

Received: 27-04-2018 Accepted: 05-10-2018

TABLE 6. Matrix of rotated components, according to varimax.

Criterion	Components	
	1	2
Frame	0,206	0,421
Topic	**0,715**	0,064
Plot	**0,711**	0,174
Creativity	**0,739**	0,200
Sentences	0,356	**0,600**
Vocabulary	0,012	**0,755**
Spelling	0,083	**0,697**

Prepared by the authors

The values in bold for each criterion correspond to a higher saturation in each of the components. In order to connect a factor with a component, it should be advisable for the first of them to have a significant factorial weight (higher than 0,50). If applying a laxer criterion (higher than 0,2), the Frame criterion could be considered for both components. Only the first was included, since it was more coherent with its construct's theoretical validity (despite having a slightly smaller weight factor).

The seven remaining factors, grouped in two components, explain 49% of the total variance, the first of those having more weight, explaining 34%. According to the structure of the instrument, the components seem therefore coherent with the Organization and Content dimension (component 1), as well as the dimension related to the grammar, lexical and orthographic aspects (component 2).

Criterion validity

The correlation between the sum of the scores obtained from the assessment of each text in the story assessment rubric (RER), and those for the standardized test PROESC, was very high r=0,79 (p<0,01). However, the teachers gave lower scores to the same texts (r=0,41; p<0,01). Nonetheless, this last correlation was higher than what we obtained between those scores and the results from the PROESC (r=0,30; p<0,01).

Specifically, table 7 shows the correlation between each of those criteria from the PROESC assessing similar aspects (the criterion "spelling" from the rubric was not included, since there are no similar criteria in the PROESC).

TABLE 7. Pearson correlation (r) between the specific criteria of both instruments.

RER	PROESC	r
Frame	Where and when Characters	0,8* 0,27*
Topic or initial event	Event with consequences	0,28*
Plot	Event with consequences Coherent ending Logical continuity Sense of unity	0,29* 0,57* 0,43* 0,46*
Creativity	Creativity Literary figures	0,34* 0,14
Sentences	Complex sentences	0,28*
Vocabulary	Vocabulary	0,48*
Spelling	-	-

Note: Statistically significant correlations: (*) $p < 0,01$. Prepared by the authors.

Internal consistency

The four elements comprising the *Organization and Content* dimension were analyzed for homogeneity, and a score of 0,57 was obtained in Cronbach's Alpha, while for the *grammar, lexical and orthographic aspects* dimension, the result was 0,60. It is a satisfactory value, higher than what we obtained with the PROESC (0,37 in both dimensions). We also observed that if we eliminate a criterion from the rubric, the Alpha value does not improve.

Received: 27-04-2018 Accepted: 05-10-2018

Interjudge reliability

To calculate the reliability in the implementation of the rubric, the degree of agreement between three evaluators was analyzed over a sample of 30 randomly chosen compositions. This was done after various training sessions for its implementation. The reliability indices obtained from Cohen's kappa method were higher than 0,85 ($p<0,01$) for all criteria. Again, the results were slightly higher than those obtained with the PROESC.

Conclusions

The formative assessment of writing is a common classroom practice. It relies largely on the subjective opinions of teachers to determine which aspects demonstrate good writing (Barton, Hamilton, & Ivanic, 2000). The literature review we did in the introduction of the present study shows the lack of objective instruments with sufficient validity and reliability. These instruments should also be practical and easily available for teachers. The ordinal scales of measurement in descriptive format (i.e. rubric) are well adapted to this objective: on one side because they offer concrete indications about the students' level of achievement within the specific criteria of a skill or task assessment, as well as what they can improve next in their learning progression. On the other side its functional and intuitive design (usually on a single sheet) facilitates its use in the classroom, not only for teachers, but for students as well.

To this effect, the main objective of this study was to design and validate a story assessment rubric (RER) for the first years of compulsory education.

The data provided in the Results section support the rubric's content validity, which suggests that the dimensional structure, operationalization of criteria, and levels of achievements are relevant and representative of the construct for its evaluation in the three levels of schooling selected. The experts (researchers and education professionals) who had the opportunity of reviewing and using the story assessment rubric (RER), valued its construction and usefulness very positively. The factor analysis of the data gathered from its application to a large sample of texts, confirmed the adequacy of the structure with two factors or dimensions

containing all the assessment criteria. The application of the rubric to the assessment of texts written by students of different ages also allowed us to confirm that the instrument satisfactorily discriminates the progressive maturation in narrative skills, expected in students of 9 to14 years of age. In addition, the global correlation of 0,79 with a standardized test (the PROESC) can be considered as an adequate *criterion validity* indicator (Roblyer & Wiencke, 2003). Another important validity indicator is the correlation with the grades that the teachers gave to these texts (Gearhart *et al.*, 1995), which was even higher than what was registered with the PROESC. We also obtained positive results from various evaluators regarding the instrument's reliability, in terms of internal consistency and the objectivity of its application. As regards consistency (measured by the Cronbach's Alpha test), the results are within the parameters recommended by many researchers in their review of research on rubrics (Graham & Perin, 2007; Jonsson & Svingby, 2007; Knoch, 2009). When applying Cohen's kappa coefficient to our rubric (higher than 0,85 for all criteria), the results were also better than for most of those studies.

In conclusion, the rubric in this study is a simple, valid, and reliable instrument for the assessment of texts written by students of 9-14 years of age. Given its characteristics and our previous experience (Montanero *et al.*, 2014), we can assume it has the potential to support formative assessment and co-assessment in the classroom.

In general, the difference of quality found between the texts of different levels of education was in line with our expectations. However, we must point out the lack of students in the last years of compulsory education who are able to write complex and well-articulated plots with truly creative elements. In particular, there is a weak progression in the rhetorical skills for most students in the first years of secondary education, a fact that was pointed out by previous studies in our education system (Artiles & Jiménez, 2007; Salvador, 2000). A possible explanation for this result is that the story assessment rubric (RER) responds to the classic structure of a simple and linear story. It does not identify a variety of resources (e.g. the analepsis, the contrast of characters in different times, etc.) that more advanced writers might use to enrich the story; this limits its discriminative capacity for such a wide range of ages. In future studies, we would like to adapt and validate different versions of the RER for an age range of only two years.

Received: 27-04-2018 Accepted: 05-10-2018

In any event, the incorporation of instruments facilitating the formative assessment of this competence in the classroom could contribute to its improvement. To confirm the usefulness of this instrument, we plan to design a longitudinal research in order to analyze the evolution of quality in the texts of students participating in assessment activities with this rubric, compared with others who do not.

Bibliographical references

Artiles, C., & Jiménez, J. E. (2007). *Proescri Primaria. Prueba de evaluación de procesos cognitivos en la escritura*. Canarias: Consejería de Educación, Cultura y Deportes del Gobierno de Canarias. Dirección General de Ordenación e Innovación Educativa.

Baker, E.L., Abedi, J.; Linn, R. L., & Niemi, D. (1995). Dimensionality and generalizability of domain-independent performance assessments. *Journal of Educational Research* 89 (4), 197–205.

Barton, D., Hamilton, M., & Ivanic, R. (2000). *Situated Literacies: reading and writing in context*. Routledge: London.

Benítez, R., & García, G. (2010). El razonamiento analógico verbal: una habilidad cognitiva esencial de la producción escrita. *Onomázein,* 22 (2), 165-194.-

Berninger, V.W, Yates, C., Cartwright, A., Rutberg, J., Remy, E., & Abbott, R. (1992). Lower-level developmental skills in beginning writing. *Reading and Writing: An interdisciplinary Journal,* 4, 257-280.

Botvin, G. J., & Sutton-Smith, B. (1977). The development of structural complexity in children's fantasy narratives. *Developmental Psychology*, 13 (4), 377-388.

Cremin, T., & Myhill, D. (2012). *Writing Voices: Creating communities of writers*. London: Routledge.

Cuetos, F., Ramos, J. L., & Ruano, E. (2002). *PROESC: Batería de evaluación de los procesos de escritura*. Madrid: TEA Ediciones.

Dunsmuir, S. Kyriacou, M. Batuwitage, S. Hinson, E. Ingram, V., & O'Sullivan, S. (2015). An evaluation of the Writing Assessment Measure (WAM) for children's narrative writing. *Assessing Writing,* 23 (2015) 1–18.

East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing*, 14, 88-115. DOI: 10.1016/j.asw.2009.04.001

Fang, Z., & Wang, Z. (2011). Beyond rubrics: Using functional language analysis to evaluate students writing. Australian Journal of Language and Literacy, 34 (2), 147-165.

Fernández, M. J., Lucero, M., & Montanero, M. (2016). Rojo sobre negro. ¿Cómo evalúan los maestros las redacciones de sus estudiantes? *Revista de educación*, 372 (abril-junio), 63-82.

Fernández, M. J. (2017). Co-evaluación de textos narrativos en la Educación Obligatoria. Universidad de Extremadura, Badajoz, España.

Ferris, D. R., & Hedgcock, J. (2014). *Teaching ESL composition: Purposes, process, and practice* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Gallavan, N., G. & Kottler, E. (2009). Constructing rubrics and assessing progress collaboratively with social studies students, social studies students. *The Social Studies*, 100 (4), 154-158.

Galve, J. (2005). BECOLE. *Batería de Evaluación Cognitiva de la Lectura y Escritura*. Instituto de Orientación Psicológica. Madrid: Editorial EOS.

García, J. N., & Marbán, J. M. (2001): Instrumento de Evaluación Individual de los Procesos Cognitivos de la Escritura (IEPCE). En J. N. García, *Dificultades de aprendizaje e intervención psicopedagógica* (pp. 141-149). Barcelona: Ariel.

Gearhart, M., Herman, J. L., Novak, J. R., & Wolf, S. A. (1995). Toward the instructional utility of large-scale writing assessment: Validation of a new narrative rubric. *Assessing Writing*, 2, 207–242.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99 (3), 445–476.

Graham, S., Harris, K.R., & Fink-Chorzempa, B. (2002). Contributions of spelling instruction to the spelling, writing, and reading of poor spellers. *Journal of Educational Psychology*, 94, 669-686.

Iandolo, G., Esposito, G., & Venuti, P. (2012). Cohesión, micro-organización, estructura narrativa y competencias verbales entre tres y once años: el desarrollo narrativo formal. *Estudios de Psicología*, 34(2), 141-160.

Janssen, G., Meier, V., & Trace, J. (2015). Building a better rubric: Mixed methods rubric revision. *Assessing Writing*, 26, 51-66. DOI: 10.1016/j.asw.2015.07.002

Received: 27-04-2018 Accepted: 05-10-2018

Jonsson A., & Svingby, G. (2007). The Use Of Scoring Rubrics: Reliability, Validity And Educational Consequences, *Educational Research Review*, 2, 130–144. DOI: 10.1016/j.edurev.2007.05.002

Kernan, Keith T. (1977). Semantic and expressive elaboration in children's narratives. *Child discourse,* 91-102.

Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275-304. DOI: 10.1177/0265532208101008

Knoch, U., Read, J., & Von Randow, J. (2007). Re-training writing raters online: How does it compare with face-to-face training? *Assessing Writing*, 12(1), 26–43. DOI: 10.1016/j.asw.2007.04.001

Kohn, A. (2006). The trouble with rubrics. *English Journal*, 95(4), 12–15.

Li, J., & Lindsey, P. (2015). Understanding variations between student and teacher application of rubrics. *Assessing Writing*, 26, 67–79. DOI: 10.1016/j.asw.2015.07.003

Lucero, M., Fernández, M. J., & Montanero, M (2018). Assessment of short stories in Elementary and Secondary Education. A study of teachers' responses to students' texts. *Studies in Educational Evaluation*, 59, 158-167.

Montanero, M., Lucero, M., & Fernández, M. J. (2014). Iterative co-evaluation with a rubric of narrative texts in Primary Education. *Infancia y Aprendizaje: Journal for the Study of Education and Development*, 37 (1), 184-220. DOI: 10.1080/02103702.2014.881653

Ochoa-Angrino, S., Aragón, L., Correa, M., & Mosquera, S. (2008). Funcionamiento metacognitivo de niños escolares en la escritura de un texto narrativo antes y después de una pauta de corrección conjunta [Versión electrónica], *Acta Colombiana de Psicología*, 11 (2), 77-88.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9(0), 129-144. DOI: http://dx.doi.org/10.1016/j.edurev.2013.01.002

Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2012). Rubrics and self assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences* (0). DOI: 10.1016/j.lindif. 2012.04.007

Roblyer, M. D., & Wiencke, W. R. (2003). Design and use of a rubric to assess and encourage interactive qualities in distance courses. *American Journal of Distance Education*, 17, 77–99.

Ryan, M. (2014). Reflexive writers; Re-thinking writing development and assessment in schools. Assessing Writing, 22, 60-74. DOI:10.1016/j.asw.2014.08.002

Salvador, F. (2000). Habilidades narrativas de alumnos de Educación Primaria en la producción de textos escritos. *Enseñanza*, 17-18, 145-163.

Schenck, A. D., & Daly, E. (2012). Building a better mousetrap: Replacing subjective writing rubrics with more empirically-sound alternatives for EFL learners. *Creative Education*, 3 (8), 1310-1325. DOI: 10.4236/ce.2012.38193

Spandel, V. (2006). Speaking my mind: In defense of rubrics. *English Journal*, 96 (1), 19-22.

Stadler, M. A., & Ward, G. C. (2005). Supporting the narrative development of young children. *Early Childhood Education Journal*, 33, 2, 73-80.

Stein, N., & Glenn, C. (1979). An analysis of story comprehension in elementary school children. En R. O. Freedle (Ed.), *Advances in discourse processing*. Vol. 2. Norwood, Nueva Jersey: Ablex.

Sundeen, T. H. (2014). Instructional rubrics: Effects of presentation options on writing quality. *Assessing Writing*, 21, 74-88. DOI: 10.1016/j.asw.2014.03.003

Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 9, 77-110.

Toro, J., & Cervera, M. (1984). *TALE. Test de análisis de la lectoescritura*. Madrid: Aprendizaje Visor.

Toro, J., Cervera, M., & Urío, C. (2002). *Escalas Magallanes de Lectura y Escritura. TALE-2000.* Vizcaya: Albor-COHS.

Yan, C. M. W., McBride-Chang, C., Wagner, R. K., Zhang, J., Wong, A. M. Y., & Shu, H. (2012). Writing quality in Chinese children: speed and fluency matter. *Reading and Writing*, 25 (7), 1499-1521.

Contact address: María Jesús Fernández Sánchez, Universidad de Extremadura, Facultad de Educación, Departamento de Ciencias de la Educación, Avda. de Elvas s/n. 06006 Badajoz, España. E-mail: mafernandezs@unex.es

Received: 27-04-2018 Accepted: 05-10-2018

Education journals worldwide: an analysis of the publications included in the 2016 Journal Citation Reports (JCR)

Las revistas de educación a nivel mundial: un análisis de las publicaciones incluidas en el Journal Citation Reports (JCR) del 2016

DOI: 10.4438/1988-592X-RE-2019-383-403

Julia Haba-Osca
Francisco González-Sala
Universitat de València
Julia Osca-Lluch
INGENIO (CSIC-UPV)

Abstract

Introduction: The evaluation of scientific activity in Spain grants personal and social recognition to professors and researchers and conditions their academic career. Articles published in journals included in the Journal Citation Reports (JCR), are the most necessary to achieve personal accreditation and sexennial research recognition. In this paper, the situation of the education journals currently included in both JCR (SCIE and SSCI) are analysed. Methodology: the sample contained 340 journals included in the Education & Educational Research, Special, Education and Psychology Educational subject categories in the 2016 JCR Science Citation Index Expanded (SCIE) edition. The distribution by countries, quartiles, language of publication and publishing institution is studied. Results: Throughout the last ten years, the number of education journals has increased in the JCR databases in all of the subject categories except in Education, Special. Only 8 countries have education journals in privileged positions (quartiles 1 and 2). The United States of America, the United Kingdom, the Netherlands and Spain are the four countries with the highest number of journals in quartiles 1 and 2. Regarding the publication language, English and Spanish are the most used

Received: 18-01-2017 Accepted: 28-09-2018

languages. It is also observed that a great number of Spanish journals publish both in English and Spanish. The three Spanish journals with greater impact factor are: Comunicar, Revista de Educación and Revista de Psicodidáctica, all publish in English and Spanish. Conclusions: Therefore it is concluded, that in the case of Spanish educational journals, the publication of the works both in English and Spanish language contributes to a greater diffusion of the papers among the national and international scientific community, increasing its visibility and impact factor.

Key words: education journals, impact factor, bibliometrics, evaluation of scientific activities, JCR

Resumen

Introducción: La evaluación de la actividad investigadora en España, otorga reconocimiento personal y social a profesores e investigadores y condiciona su carrera académica. Los artículos publicados en revistas incluidas en el Journal Citation Reports (JCR), son los más necesarios para conseguir la acreditación personal y sexenios de investigación. En este trabajo, se analiza la situación de las revistas de educación incluidas actualmente en los JCR (SCIE y SSCI). Metodología: la muestra la conformaron 340 revistas incluidas en las categorías temáticas *Education & Educational Research*, *Special, Education* y *Psychology Educational* de la edición del JCR del Science Citation Index Expanded (SCIE) del 2016. Se estudia la distribución por países, cuartiles, idioma de publicación e institución editora. Se estudia la distribución por países, cuartiles, idioma de publicación e institución editora. Resultados: A lo largo de los últimos diez años, el número de revistas de educación ha ido aumentando en las bases de datos JCR en todas las categorías temáticas excepto en la categoría *Education, Special*. Solamente 8 países tienen revistas de educación situadas en posiciones de privilegio (cuartiles 1 y 2). Estados Unidos, Reino Unido, Holanda y España, son los cuatro países con mayor número de revistas en los cuartiles 1 y 2. En cuanto al idioma de publicación, el inglés y el español son los idiomas más utilizados. Se observa que un gran número de revistas españolas publican sus trabajos en inglés y en español. Las tres revistas españolas con mayor factor de impacto: *Comunicar*, *Revista de Educación* y *Revista de Psicodidáctica*, publican sus trabajos en inglés y español. Conclusiones: Se concluye que, en el caso de las revistas españolas de educación, la publicación de los trabajos en inglés y español, contribuye a una mayor difusión de los mismos entre la comunidad científica nacional e internacional, aumentando su visibilidad y factor de impacto.

Palabras clave: revistas de educación, factor de impacto, bibliometría, evaluación de la actividad científica, JCR.

Introduction

Scientific journals are the main tools available to researchers today to disseminate the results of their research. This means must be permanent, punctual, rigorous and transparent, becoming the official and public record of knowledge. But, in addition, periodical publications play an important role in the process of scientific communication, since they constitute the main way for the validation of new knowledge and make possible its dissemination within the community of researchers, and from this to society and its agents.

Scientific journals have a fundamental role in the different stages of research activity, since they are the way in which researchers obtain recognition for their contributions to scientific progress. The publication of a work in a recognized prestigious journal can contribute to increase the personal and social recognition of professors and researchers (Carbonell and Calvó, 2009), also directly conditioning the progression in their academic career (Jiménez-Contreras, Torres-Salinas, Ruiz-Pérez and Delgado, 2010, Ruiz-Pérez, Martín-Martín and Delgado, 2015) and increase their economic rewards. In the current context of globalization, international visibility has also become a fundamental asset for higher education institutions, given the need to strengthen ties with other institutions to optimize resources and complement capacities and as a strategy for obtaining resources and positioning in the global scope. Being visible internationally contributes to attracting prestigious students, professors and researchers and being able to link with other partners in order to request projects in international calls (De Filippo, Pandiella-Dominique and Sanz-Casado, 2017), visibility brings with it recognition and resources (Docampo, 2008).

Bibliometric studies on the production and impact of scientific publications are the main means used by scientists to know the quality of their research, as well as the greater or lesser relevance of the institutions to which they are linked (Cronin, 1984). The evaluation of scientific production and of science in general has traditionally been limited and focused on scientific journals (Giménez-Toledo, Mañana-Rodríguez and Tejada-Artigas, 2015). However, work on the consumption of information has put the existence of marked and evident differences between publication and citation habits among the scientists of the different disciplines (Archambault and Larivière, 2010, Osca-Lluch, Veyrat and

Received: 18-01-2017 Accepted: 28-09-2018

Morales, 2013, Dorta-González and Ramirez-Sánchez, 2014). One of the aspects that most differentiates some disciplines from others is that related to the vehicle for disseminating research results. All the studies based on citation computations coincide in stating that in the experimental sciences the information is transmitted mainly through journal articles, which add percentages of citations higher than 80 percent, followed by books far away, with around 10 per cent. On the contrary, in social sciences and humanities and, for various reasons, also in practical applications, books predominate, accounting for between 50 and 65 percent of citations, while journal articles have a relatively modest weight, which manifests itself in percentages between 10 and 35 (Broadus, 1971). The relevance of books is indisputable and the absence of exhaustive international databases that cover the elements and information necessary for the evaluation of this type of publications has prompted several European countries to develop customized information systems for the registration of academic books (Giménez-Toledo et al., 2016, Giménez-toledo, Mañana-Rodriguez and Sivertsen, 2017). This different behavior at the time of publication has, in general, its corresponding reflection in the citation guidelines of the authors. From the work done on the humanists, common characteristics have been obtained to these researchers, among which it is worth mentioning that it is observed that they preferably use formal channels to update their knowledge, and within them, they usually consult more monographs than periodicals. Informal channels are very important, although they are used mainly to maintain personal contacts with colleagues, although they tend to work isolated, so there are relatively few collaboration groups in this area and, in general, they have a great language capacity, so they can make use of documents written in different languages (Osca-Lluch, Veyrat and Morales, 2013). However, the evaluation of scientific production and of science in general has traditionally been limited and focused on scientific journals (Giménez-Toledo, Mañana-Rodríguez and Tejada-Artigas, 2015), and unlike what happens with journals, there are few instruments and indicators that exist to evaluate the quality of the monographs (Giménez-Toledo and Tejada-Artigas, 2015).

In recent years there have been major changes in terms of the edition, evaluation and hierarchization of the merit of journals (Santos and Fernández-Ríos, 2016). The citation of articles is the most important indicator of visibility when measuring the academic impact of journals

Received: 18-01-2017 Accepted: 28-09-2018

and the knowledge published there. Impact indicators are values that are calculated from the number of citations recorded by the works published in a journal, which are generally accepted as measures that indicate the importance, impact, prestige or visibility of a journal with respect to to the others in the same scientific field and that serve to establish rankings of journals indexed in the databases themselves, according to indicators that they themselves calculate.

The most recognized and frequently used indicator to evaluate the quality of scientific publications and measure the impact of a journal is the impact factor of the Journal Citation Reports (JCR), a bibliometric indicator currently of the Clarivate Analytics company, which reflects the average of citation of articles published in a scientific journal indexed in the Web of Science system in the Science Citation Index Expanded indexes (SCIE) or in the Social Sciences Citation Index (SSCI), not in the Arts & Humanities Citation Index (A&HCI). The use of the impact factor has been widely criticized, as being a limitation for non-English publications (González-Alcaide, Valderrama-Zurián and Aleixandre-Benavent, 2012) or because the number of citations does not measure the quality of the article (Velasco, Eiros, Pinilla and San Román, 2012). Over time, the impact indicators have encountered some drawbacks, which has led to the formulation of new ones that always try to be a reliable measure of the quality, prestige or importance of the publications (Bormann and Daniel, 2007, Buela-Casal, 2002, 2003, Dodson, De Souza and Dos Santos, 2012, Egghe, 2006, Hirsch, 2010, Hirsch and Buela-Casal, 2014, Ramirez Martinez, Martinez Ruiz and Castellanos, 2012). However, the new indicators have not had luck or as much acceptance as the impact factor (IF) and are not exempt from criticism either. One of the most criticized aspects of the use of the impact factor of the JCR is the scarce presence of publications from some non-English speaking countries, despite the fact that the most developed Ibero-American countries have carried out rigorous policies to promote the highest quality national publications. (García-Pereira and Quevedo-Blasco, 2015, Quevedo-Blasco and López-López, 2011, Rodriguez Yunta, 2010).

The implantation in Spain of a system of evaluation of the research activity has been very controversial from its origins, especially since it gave rise to economic incentives through the known research six-year period - *sexenios* in Spanish - (Galán, 2014). The system grants personal and social recognition to professors and researchers, also conditioning

directly the progression in their academic career. Teachers who fail to be recognized their research are forced to spend most of their day in teaching, so they have no time to investigate (Ruíz-Pérez, Martín-Martín and Delgado, 2015).

The recently published criteria that indicate the merits necessary to achieve the evaluation of the research activity reiterate that "articles published in journals of recognized value will be preferentially assessed, accepting as such those included in the lists by scientific areas of the Journal Citation Reports (Social Sciences Edition) and Journal Citation Reports (Science Edition) of the "Web of Science" and among other considerations, it is also indicated that "the repeated publication of works in journals or publishers belonging or associated to the same body where the applicant will be unfavorably evaluated. performs its research, as well as the different contributions are duplicated or result iterative without effective innovation. It will also be unfavorably assessed that more than one contribution forms part of the same book or journal number. Likewise, when the work shows a high level of self-citations" (CNEAI, 2017).

For all these reasons, knowing the journals of a particular scientific area or discipline is essential for researchers, so that they can select which journals are the most suitable to publish their work and, in recent years, they have made different studies whose objective is the analysis of the diffusion and position of scientific journals in different databases (Castillo and Carretón, 2010, García-Pereira and Quevedo-Blasco, 2015, González-Sala and Osca-Lluch, 2016, Olivas-Avila, Musi-Lechuga, Quevedo-Blasco and Luna-Hernández, 2012, Quevedo-Blasco, 2013, Quevedo-Blasco and López-López, 2011, Repiso, Jiménez-Contreras and Aguaded, 2017).

If the journals, as means of publication and dissemination of the research, have a great relevance, the platforms or databases where they are indexed are no less important. The general objective of this work is to characterize the education journals collected in the Journal Citation Reports (JCR). The analysis focuses on two fundamental aspects, the position of the journals and distribution according to the different countries, and the editorial aspects with the purpose of knowing the journals of excellence in this scientific discipline.

Received: 18-01-2017 Accepted: 28-09-2018

Method

Analysis units and materials

All the journals indexed in the thematic categories "Education & Educational Research", "Special Education" and "Psychology Educational" of the edition of the Journal Citation Reports (JCR) of the Social Sciences Citation Index (SSCI) and of the thematic category "Education, Scientific Disciplines" of the edition were analysed in the 2016 Journal Citation Reports (JCR) of the Science Citation Index Expanded (SCIE) (corresponding to the year 2017)[1].

In the JCR databases (SCIE and SSCI), the same journal can be included in one or both of the JCR databases and, in addition, it can also be indexed in one or several thematic categories in each of the databases. Of the four categories studied: "Education & Educational Research", is the one with the largest number of journals related to the whole range of works on education, from theoretical to applied. It indexes resources on pedagogy and methodology, as well as on the history of education, reading, curricular studies, educational policy and sociology and economics in education, as well as on the use of computers in the classroom. The "Education, Special" category includes journals that publish works related to the education and development of people with special needs, including the gifted and those with learning disabilities. "Psychology, Educational" includes publications on educational psychology, creative behavior, education sciences, reading research and school psychology. Finally, the category "Education, Scientific Disciplines", includes journals that cover all the educational resources in the different scientific disciplines.

All the journals included in the different categories related to education have been reviewed, in order to detect the existing overlap between them. Once the duplicated journals have been eliminated, that is, those that have been classified in more than one thematic category, a total of 340 journals have been identified.

(1) The data for this study were collected in January 2018.

Received: 18-01-2017 Accepted: 28-09-2018

Procedure

The distribution by thematic categories, countries, institutional typology, language and quartiles occupied in the JCR database was analyzed from each of the journals[2].

Results

Journals related to the discipline of Education, indexed in the 2016 JCR database can be classified thematically, in one or more of the four categories mentioned above. The category "Education & Educational Research" is the one that currently includes a greater number of journals (236). The second place, in terms of the number of journals collected, is the "Psychology Educational" category, with 58 journals. The third and fourth place are occupied by the "Education, Scientific Disciplines" and "Education, Special" categories, with 41 and 38 journals respectively.

It should be noted that, over the last few years, the number of education journals has been increasing in the JCR databases in all the thematic categories except in the "Education, Special" category, which drops slightly in the last JCR edition, that of 2016, corresponding to the 2017 edition, as shown in Table I, and, on the other hand, the category with the highest growth is "Education & Educational Research". When it is observed that the average impact factor of the different thematic categories throughout the period analyzed, it is depicted that the highest factors do not always coincide with the categories that have a greater number of journals. It is also striking that within the same category there is no equivalence between the greater number of indexed journals and the greater impact factor of that category. In the categories "Education, Scientific Disciplines", "Education, Special" and "Psychology, Educational" is in 2013 and 2014 when values of higher impact factor are obtained in each of these categories, while in the case of "Education & Educational Research", are 2014 and 2016.

(2) All the information related to the journals included in the JCR databases (SCIE and SSCI) can be downloaded from ISI Web of Knowledge in PDF, CVS or XLS format, at https://jcr.incites.thomsonreuters.com/

Received: 18-01-2017 Accepted: 28-09-2018

TABLE I. Distribution of the number of journals in JCR per year, thematic category and impact factor (IF) average of the thematic category.

Years	Education & Educational Research	IF	Education, Scientific Disciplines	IF	Education, Special	IF	Psychology Educational	IF
2007	105	0,548	23	0,663	29	0,529	38	0,813
2008	113	0,711	24	0,883	30	1,000	42	1,037
2009	139	0,723	27	0,779	30	1,117	44	0,982
2010	184	0,906	33	1,220	36	1,373	50	1,416
2011	206	0,708	33	0,902	37	0,906	51	1,116
2012	219	0,644	34	0,945	36	0,914	51	1,100
2013	219	0,914	36	1,430	37	1,694	53	1,580
2014	224	0,922	37	1,352	39	1,447	55	1,687
2015	231	0,896	40	1,211	39	1,000	57	1,219
2016	236	1,107	41	1,329	38	1,168	58	1,406

The same journal can be classified into different thematic categories and, in addition, occupy different positions and quartiles in each of the categories in which it has been included. When the number of education journals is analyzed, eliminating the overlaps between the four analyzed categories indexed in the JCR 2016, it is observed that the number of publications is 340.

The analysis of the distribution by countries shows that the education journals included in the JCR are edited by a total of 22 countries. The United States and the United Kingdom are the two countries with 78.66% of the education journals included in the 2016 JCR. There are only 6 countries that have more than 5 education journals currently indexed in the JCR, which are United States (156 journals), United Kingdom (111 journals), Netherlands (22 journals), Spain (9 journals), Germany (8 journals) and Australia (8 journals).

The analysis of the distribution of the journals by quartiles (see table II) shows that only six countries have a journal located in the quartile of greatest impact or first quartile (Canada, Netherlands New Zealand, Spain, United Kingdom and United States). The number of countries with journals included in Quartile 2 is seven (Australia, Netherlands, New Zealand, Portugal, Spain, United Kingdom and United States). Therefore,

Received: 18-01-2017 Accepted: 28-09-2018

we can see that of the 22 countries with education journals indexed in JCR, there are only 8 countries with journals that are included in privileged positions (quartiles 1 and 2). As it can be seen, the largest number of countries has their journals ranked in quartile 4, except Belgium, Canada, Denmark, Mexico and Portugal.

TABLE II. Distribution of journals by country and quartile *

COUNTRIES	Q1	Q2	Q3	Q4	Total journals
Australia		1	3	4	8
Belgium			1		1
Brazil				1	1
Canada	1		1		2
Croatia				1	1
Denmark			1		1
Germany			3	5	8
India				1	1
Italy				1	1
Lithuania				1	1
Mexico			1		1
Netherlands	6	9	4	3	18
New Zealand	1	1	1	1	4
Philippines				1	1
Portugal		1			1
Russia				1	1
South Africa			2	2	4
South Korea				2	2
Spain	1	2	2	4	9
Turkey			1	2	3
United Kingdom	35	20	31	25	111
United States	53	42	39	22	157
Total	**97**	**76**	**90**	**77**	**340**

*2016 JCR data. Only one journal is shown per quartile. When the same journal is included in more than one thematic category, only the quartile in which it is best positioned has been counted in the table.

Received: 18-01-2017 Accepted: 28-09-2018

If we look at the Spanish education journals included in the JCR databases (SSCI and SCIE), we can see that the first Spanish scientific journals of this discipline did not appear until the 20th century, specifically in the 40s. We had to wait until 1980 to find a significant increase in the scarce periodic publications focused on education. Two journals stand out as pioneers: *Revista Nacional de Educación* (1941), which later adopted the title of *Revista de Educación* (1952). Since its inception it has been published by the Ministry of Education and is currently a recognized journal in the entire academic and scientific community (Ruiz-Corbella, Galán and Diestro, 2014). Since 2010, it has been indexed in the JCR database of the SSCI, in the "Education & Educational Research" thematic area. The second is *Revista Española de Pedagogía* (1943), promoted by the San José de Calasanz Institute of Pedagogy of the Higher Council for Scientific Research (CSIC - *Consejo Superior de Investigaciones Científicas*, in Spanish), and which is currently edited by the *Universidad Internacional de La Rioja* (Ruiz-Corbella, Galán and Diestro, 2014). This publication is also included in the "Education & Educational Research" thematic area of the JCR database (SSCI) since 2007. The other Spanish journals of this discipline that are currently included in the JCR are:

- *Infancia y Aprendizaje* (1978), edited by *Fundación Infancia y Aprendizaje* and Taylor & Francis, and which is indexed in the categories "Psychology", "Educational" and "Psychology, Developmental" JCR (SSCI) since 2008.
- *Enseñanza de las Ciencias* (1983), of the *Universitat Autònoma de Barcelona*, indexed in the "Education & Educational Research" category of the JCR (SSCI) since 2010.
- *Comunicar* (1994), edited by the Comunicar Group, which has been indexed since 2009 in the thematic categories "Education & Educational Research" and "Communication" of the JCR (SSCI).
- *Cultura y Educación* (1996), edited by *Fundación Infancia y Aprendizaje* and Taylor & Francis. It has been included since 2010 in the "Education & Educational Research" category of the JCR (SSCI).
- *Revista de Psicodidáctica* (1996), edited by *Universidad del País Vasco*. Indexed in the "Psychology, Educational" category of the JCR (SSCI) since 2009.

- *Educación XX1* (1998), published by Universidad Nacional de Educación a Distancia (UNED), indexed in the JCR (SSCI) "Education & Educational Research" category since 2011.
- *Porta Linguarum* (2004), edited by *Universidad de Granada*. It is indexed in the thematic categories "Education & Educational Research" and "Linguistics" of the JCR (SSCI) since 2010.

When the language of the publications is analyzed, it is observed that the works published in the education journals can be published in 6 different languages. The most used language by publications is English (94.41%, 321 journals), the second place is occupied by journals that admit works written in more than one language (multilanguage), with 3.82% (13 journals). In the case of publications that publish their works in more than one language, the most frequent cases are those that accept works in English and Spanish or English and German. The other 5 languages used by the publications are Spanish (0.59%, 2 journals), Croatian, Italian, Russian and Turkish, with a percentage of 0.29 (1 journal) each.

There are four groups of institutions publish the journals of education indexed in the JCR: Commercial publishing houses, who publish 77.94% of the journals; Foundations and Associations (11.47%); Universities (10.29%) and Government Organizations (0.29%). Within the group of commercial publishers the most outstanding include Routledge Journals, Taylor & Francis LTD (98 journals), Wiley-Blackwell (75 journals), Sage Publications LTD (42 journals), Springer (31 journals) and Elsevier (21 journals).

Regarding the impact factor of the journals, it can be seen that the journals included in the categories that index a greater number of journals –"Education & Educational Research" and "Psychology Educational"– have the greatest impact factor. On the contrary, journals that have a lower impact factor, are those included in the categories that have fewer journals, such as "Education Special" and "Education, Scientific Disciplines".

The journal that has the greatest impact factor is *Education Psychologist* (IF 6,257), which ranks first in the "Education & Educational Research" and "Psychology Educational" categories. The second journal with the highest impact factor is *Review of Educational Research* (IF 5,263), which is indexed in the "Education & Educational Research" category. In the "Education Scientific Disciplines" category, the journal with the highest impact factor is *Academic Medicine* (IF 5,255) and in the "Education

Received: 18-01-2017 Accepted: 28-09-2018

Special" category, the journal with the highest impact factor is *Journal of Influency Disorders* (IF 2,714).

Discussion

This study aims to provide an overview of the situation of education journals in the 2016 JCR, where a total of 340 journals classified in the 2016 Journal Citation Reports (published in 2017) have been indexed. Journals play an important role in the evaluation of scientific activity, both of the authors and their institutions and, for this reason, it is fundamental that researchers know which journals are the most suitable for the publication of their work. The legislative changes introduced since the 80s in Spain made it clear that the evaluation and recognition of university teaching staff necessarily went through the evidences of their publications, especially in the form of articles of proven quality or, in other words, journals with an impact. It is undeniable that this evaluation policy has prompted Spanish journals to adapt to the patterns of international dissemination of scientific productivity to improve its impact (Ruiz-Corbella, Galán and Diestro, 2014). However, for some authors, these policies can provoke an authentic perversion of the system, since it can lead researchers to focus on different strategies to publish and be cited, to the detriment of other quality criteria and scientific ethics (Buela-Casal, 2014).

One of the indicators that due to its international relevance has become a benchmark for assessing the quality of periodicals is the Journal Citation Report (JCR). In the case of education journals included in JCR, as it was expected given the characteristics of these databases, the countries that have the largest number of journals are the United Kingdom and the United States. However, we must highlight the presence of journals belonging to countries such as the Netherlands and Spain, which occupy the third and fourth place in number of publications in this discipline.

The visibility of Spanish scientific publications has increased considerably and it is observed that, as of 2007, there has been a significant increase in the number of indexed journals (Moreno-Pulido, López-González, Rubio-Garay, Saúl and Sánchez-Elvira-Paniagua, 2013). The analysis of the education journals in JCR reveals that there are 9

Spanish education journals included in JCR and that there are only 3 currently occupying privileged positions (quartiles 1 and 2).

Regarding the language of publication, English is accepted in 329 of the 340 journals, and Spanish is the second most widely used publication language. Although a large number of Spanish journals publish their works in English and Spanish (7 journals). The three Spanish journals with the greatest impact factor, such as *Comunicar* (quartile 1 in the categories "Education & Educational Research" and "Communication"); *Revista de Educación* (quartile 2 in the "Education & Educational Research" category) and *Revista de Psicodidactica* (quartile 2 in the "Psychology Educational" category) publish their works in English and Spanish, which contributes to a greater dissemination among the national and international scientific community.

The study of the impact factor of the journals reveals that in the categories where there is a greater number of indexed journals, they are not always those that obtain a higher impact factor, but it is related to the citations that the works receive. Some studies reveal that more and more are cited, but that some elements appear that call attention to growth in the self-citation rate that surely have more to do with editorial strategies in pursuit of a higher impact index than a real increase in influence of these journals (Fernández-Quijada, 2010).

As a conclusion to this work, it can be affirmed that the implementation of an evaluation policy based on international indicators has contributed to improving the quality and visibility of Spanish education journals. Spain occupies the fourth place in the ranking of countries that have education journals located in quartiles 1 and 2 in the JCR, after the Netherlands, United Kingdom and United States, which are the three countries that occupy the first three places. Of the 9 Spanish journals currently indexed in JCR 3 of them, are located occupying privileged positions (quartiles 1 and 2), which comes to reveal the acceptance of the works published in the same among the scientific community and one of the conclusions of this work is that it is recommended, the publication of the works in both languages (English and Spanish), since it contributes to improve the dissemination and visibility of the works published also among the Spanish native-speaking community. It is observed that the Spanish educational journals have more and more presence in the JCR, sharing the results of Quevedo-Blasco and López-López (2011) and García-Pereira and Quevedo-Blasco (2015), it would be desirable that

Received: 18-01-2017 Accepted: 28-09-2018

these Spanish journals not only will increase their impact factor, but new ones will be indexed in these databases, in order to avoid leaks to multidisciplinary journals or other more general areas, whose journals have a greater impact factor.

References

Archambault, E. y Larivière, V. (2010). *The limits of bibliometrics for the analysis of the social sciences and humanities literature.* En: 2010 World Social Science Report: Knowledge Divides. Paris: UNESCO, 251-254.

Bormann, L. y Daniel, H.D. (2007). Convergent validation of peer review decisions using the h index-Extent of and reason for type I and type II errors. *Journal of Informetrics*, 1 (3), 204-213.

Broadus, R.N. (1971): The Literature of the Social Science: A Survey of Citation Studies. *International Social Science Journal*, 23, 236-243.

Buela-Casal, G. (2002). Evaluación de la investigación científica: "El criterio de la mayoría": El factor de impacto, el factor de prestigio y los "diez mandamientos para incrementar las citas". *Análisis y Modificación de Conducta*, 28, 119, 455-476.

Buela-Casal, G. (2003). Evaluación de la calidad de los artículos científicos: propuesta del factor de impacto ponderado y de un índice de calidad. *Psicothema,* 15, 1, 23-35.

Buela-Casal, G. (2014). Pathological publishing: a new psychological disorder with legal consequences? *The European Journal of Psychology Applied to Legal Context*, 6(2), 91-97.

Carbonell, X. y Calvó, N. (2009). Las revistas españolas de psicología Cómo elegir la revista donde publicar. *abonos@lesarts.com*, 25, 209-216.

Castillo, A. y Carretón, M.C. (2010). Investigación en comunicación. Estudio bibliométrico de las revistas de Comunicación en España. *Comunicación y Sociedad*, 2, 289-327.

CNEAI (2017). Resolución de 23 de noviembre de 2017, de la Comisión Nacional Evaluadora de la Actividad Investigadora, por la que se publican los criterios específicos aprobados para cada uno de los

campos de evaluación. BOE núm. 292, de 1/12/2017, p. 117060-117080.

De Filippo, D., Pandiella-Dominique, A. y Sanz-Casado, E. (2017). Indicadores para el análisis de la visibilidad internacional de las universidades españolas. *Revista de Educación*, 376, 163-199. DOI: 10.4438/1988-592X-RE-2017-376-348.

Docampo, D. (2008). Rankings internacionales y calidad de los sistemas universitarios. *Revista de Educación*, número extraordinario, 149-176.

Dodson, M.V., De Souza Duarte, M. y Dos Santos Dias, L.A. (2012). SP-index: The measure of the scientific production of researchers. *Biochemical and Biophysical Research Communications*, 425 (4), 701-702.

Dorta-González, P. y Ramirez-Sánchez, M. (2014). Producción e impacto de las instituciones españolas de investigación en Arts & Humanities Citation Index (2003-2012). *Arbor*, a191. Doi: http://dx.doi.org/10.3989/arbor.2014.770n6012

Egghe, L. (2006). Theory and practice of the g-index. *Scientometrics,* 69 (1), 131-152.

Fernández-Quijada, D. (2010). El perfil de las revistas españolas de comunicación (2007-2008). *Revista Española de Documentación Científica*, 33 (4), 553-581.

Galán, A. (2014). Evaluación de la producción científica en educación. ¿Una nueva crisis? *Bordón*, 66 (2), 7-10.

García-Pereira, S. y Quevedo-Blasco, R. (2015). Análisis de las revistas iberoamericanas de Psicología y Educación indexadas en el Journal Citation Reports del 2013. *European Journal of Education and Psychology*, 8, 85-96.

Giménez-Toledo, E., Mañana-Rodríguez, J. y Sivertsen, G. (2017). Scholarly book publishing: Its information sources for evaluation in the social and humanities. *Research Evaluation*, 26(2), 91-101.

Giménez-Toledo, E., Mañana-Rodríguez, J., Engels, T.C., Ingwersen, P., Polonen, J., Sivertsen, G., Verleysen, F.T. y Zuccala, A.A. (2016). Taking scholarly books into account. Current developments in five European countries. *Scientometrics*, 107, 685-689.

Giménez-Toledo, E., Mañana-Rodríguez, J. y Tejada-Artigas, C.M. (2015). Revisión de iniciativas nacionales e internacionales sobre evaluación de libros y editoriales. *El Profesional de la Información*, 24, 705-716.

Received: 18-01-2017 Accepted: 28-09-2018

Giménez-Toledo, E. y Tejada-Artigas, C. M. (2015). Proceso de publicación, calidad y prestigio de las editoriales científicas en educación. Educación XX1, 18(1), 17-44.

González-Sala, F. y Osca-Lluch, J. (2016). Análisis de las publicaciones españolas en la categoría Psychology Educational de la Web of Science durante el período 2004-2013. *Aula Abierta*, 44, 46-54.

Hirsch, J.E. (2010). An index to quantify an individual´s scientific research output that takes into account the effect of multiple coauthorship. *Scientometrics*, 85 (3), 741-754.

Hirsch, J.E. y Buela-Casal, G. (2014). The meaning of the h-index. *International Journal of Clinical and Health Psychology*, 14, 2, 161-164.

Jiménez-Contreras, E., Torres-Salinas, D., Ruiz-Pérez, R. y Delgado López-Cózar, E. (2010). Investigación de excelencia en España: ¿protagonistas o papeles secundarios? *Medicina Clínica*, 134 (2), 76-81.

Moreno-Pulido, A., López-González, M.A., Rubio-Garay, F., Saúl, L.A. y Sanchez-Elvira.Paniagua, A. (2013). Evolución de las revistas españolas de ciencias sociales en el Journal Citation Reports (2006-2010) y su valoración como indicio de calidad en la normativa evaluadora. *Revista Española de Documentación Científica*, 36 (3), e014. doi: http://dx.doi.org/10.3989/redc.2013.3.987

Olivas-Avila, J.A., Musi-Lechuga, B., Quevedo-Blasco, R. y Luna-Hernández, J.R. (2012). Índice de internacionalidad de las revistas iberoamericanas de psicología en el Journal Citation Report (2011). *Revista Latinoamericana de Psicología*, 44 (3), 175-183.

Osca-Lluch, J., Veyrat, A. y Morales, J. (2013). El consumo de información en humanidades. Arbor, 189,760, a026. Doi: http://dx.doi.org/10.3989/arbor.2013.760n2012

Quevedo-Blasco, R. (2013). Revistas iberoamericanas de psicología indexadas en el Journal Citation Reports de 2011. *Revista Mexicana de Psicología*, 30 (1), 1-10.

Quevedo-Blasco, R. y López-López, W. (2011). Situación de las revistas iberoamericanas de psicología en el Journal Citation Reports de 2011. *Universitas Psychologica*, 10 (3), 955-965.

Repiso, R., Jiménez-Contreras, E. y Aguaded, I. (2017). Revistas iberoamericanas de educación en Scielo Citation Index y Emerging Source Citation Index. *Revista Española de Documentación Científica*, 40 (4): e186. doi://dx.doi.org/10.3989/redc.2017.4.1445

Rodriguez-Yunta, L. (2010). Las revistas iberoamericanas en Web of Science y Scopus: visibilidad internacional e indicadores de calidad. En: J. Rios Ortega, coord. *VII Seminario Hispano-Mexicano de Investigación en Bibliotecología y Documentación* (pp.347-363). México: UNAM.

Ruiz-Corbella, M., Galán, A. y Diestro, A. (2014). Las revistas científicas de educación en España: evolución y perspectivas de futuro. *Relieve. Revista Electrónica de Investigación y Evaluación Educativa*, 20 (2). Doi: 10.7203/relieve.20.2.4361

Ruiz-Pérez, R., Martin-Martin, A. y Delgado López-Cózar, E. (2015). Las revistas universitarias en el marco de los criterios de evaluación de la actividad investigadora en España. *Revista Española de Documentación Científica*, 38 (2): e081.doi: http://dx.doi.org/10.3989/redc.2015.2.1191

Ramirez Martinez, D.C., Martinez Ruiz, L.C. y Castellanos Dominguez, O.F. (2012). *Divulgación y difusión del conocimiento: las revistas científicas*. Bogotá: Universidad Nacional de Colombia.

Santos Rego, M.A. y Fernández-Rios, L. (2016). El factor de impacto y el futuro de las revistas académicas. El riesgo de patologización. *Innovación Educativa*, 16, 35-51.

Velasco, B., Eiros, J.M., Pinilla, J.M. y San Román, J. A. (2012). La utilización de los indicadores bibliométricos para evaluar la actividad investigadora. *Aula Abierta*, 40 (2), 75-84.

Contact address: Julia Haba-Osca. Universitat de València. Facultat de Filologia, Traducció i Comunicació, Departament de Filologia Anglesa i Alemanya. Avda. Blasco Ibáñez, 21, 46010, Valencia. E-mail: julia.haba@uv.es

Received: 18-01-2017 Accepted: 28-09-2018

Assessment of pre-service teachers' Science knowledge: the case of Valencian Community in Spain

Evaluación del Conocimiento Científico en Maestros en formación inicial: el caso de la Comunidad Valenciana

DOI: 10.4438/1988-592X-RE-2019-383-404

José Javier Verdugo-Perona
Florida Universitaria
Joan Josep Solaz-Portolés
Vicente Sanjosé
Universitat de València

Abstract

Scientific literacy is one of the main objectives in compulsory education, but this literacy can hardly be achieved if primary teachers have a low level of scientific knowledge (Appleton, 2003). In Spain, some studies suggest that primary education teachers show deficiencies in their scientific training (Cañal, 2000 y 2008, Porlán, 2010). For this reason, an assessment of pre-service elementary school teachers' knowledge on basic science is developed with enough external validity (sampling error about 5%). The scope of the study is the Valencian Community in Spain. Conceptual and procedural knowledge are assessed independently before as well as after the science subjects in the specific Degree of pre-service teacher training at University. Prior studies and their branch are taken into account in some descriptions and analyses. The possible effects of the more or less classroom hours devoted to science by the three public universities in the considered Region, are also studied. Results show significant improvements in the levels of conceptual and procedural knowledge along the Degree, but of little effect size. Outcomes suggest that some teachers' educational needs are not yet sufficiently met at the considered universities.

Key words: Science Education; Science Concepts; Science Processes skills; Preservice teachers, Assessment.

Received: 03-08-2017 Accepted: 07-09-2018

Resumen

La alfabetización científica es en uno de los objetivos principales de la educación obligatoria, pero difícilmente se puede conseguir esta alfabetización si los profesores de enseñanza primaria tienen un pobre conocimiento científico (Appleton, 2003). Algunos estudios sugieren que, en España, el profesorado de educación primaria presenta carencias en su formación científica (Cañal, 2000 y 2008; Porlán, 2010). Por este motivo, se desarrolla un estudio sobre el conocimiento sobre ciencia básica de los maestros en formación en las universidades públicas de la Comunidad Valenciana, con suficiente validez externa (error de muestreo cercano a 5%). Se evalúa independientemente el conocimiento sobre conceptos básicos y sobre procesos elementales de la actividad científica, tanto antes de recibir formación como al finalizar la formación prevista en el Grado de Maestro de Primaria. La titulación de acceso al Grado y la especialidad de los estudios pre-grado, son consideradas para describir y analizar los resultados. Complementariamente, se estudia el posible efecto de las diferencias entre universidades en las horas de ciencias ofertadas. Los hallazgos muestran mejoras significativas, aunque de pequeño tamaño, en el conocimiento analizado. También evidencian necesidades formativas no suficientemente bien atendidas en el Grado universitario.

Palabras clave: Enseñanza de las ciencias; conceptos científicos; procesos de la ciencia; maestros/as en formación; evaluación.

Introduction

Scientific literacy has become in one of the main goals in compulsory education. In fact, this concept is mentioned in the last reforms of the educational laws. One of the aims of Scientific Literacy is developing citizens' critical thinking (Acevedo, 2004) especially concerning decisions to be made on daily life socio-political issues involving basic scientific knowledge (Gil & Vilches, 2006). The National Research Council (NRC) in USA defines Scientific Literacy (SL onwards) as "…the knowledge and understanding of scientific concepts and processes required for personal decision making, participation in civic and cultural affairs, and economic productivity" (NRC, 1996, p. 22). NRC suggests Inquiry as the best instructional approach to achieve these goals. In good agreement, Hurd (1998) has conceived SL as "…a civic competency required for rational thinking about science in relation to personal, social, political, economic problems, and issues that one is likely to meet throughout life (p. 410).

Received: 03-08-2017 Accepted: 07-09-2018

The basis for an appropriate citizen education should be stablished in Primary Education (Lloyd, Smith, Fay, Khang, Wah & Sai, 1998), an important period to increase people's SL. However, poor teachers' scientific knowledge (Appleton, 2003) and then, a teachers' inefficient pedagogical content knowledge in sciences have been found as main causes of low levels of SL in citizens (Kang, 2007). In fact some conceptual errors, similar to the ones in their students, have been found in different samples of Primary teachers (Krugar, Summers & Palacio, 1990; Harlen & Holroyd, 1997; Papageorgiou, Stamovlasis & Johnson, 2010; Summers & Krugar, 1992). Probably, ineffective pre- and in-service teacher education programmes are in the origin of these problems, causing teachers' insecurity and lack of confidence when teaching children (Jarvis & Pell, 2004; García-Ruiz & Sánchez, 2006; Murphy & Smith, 2012). Other possible consequences of this poor teaching are children's negative attitudes towards science (Harlen & Holroyd, 1997; Jurievi, Glaar, Puko & Devetak, 2008).

Hence, there is some evidence suggesting that improving citizens' SL requires a deep revision of the teacher education programmes, in particular those of elementary pre-service teachers. One of the focuses of the revision should be to ensure higher enough levels of domain in elementary sciences.

In the Spanish context the lack of such domain in a variety of scientific subjects has been pointed out by Cañal (2000; 2008) and Porlán (2010). These authors highlighted the impossibility of constructing the appropriate Pedagogical Content Knowledge necessary to manage successful processes of learning and teaching in the classroom.

Despite the importance of the lack of scientific domain, most studies involving pre- and in-service Spanish teachers focused mainly on their epistemic conceptions on how Science is elaborated and validated, i.e. on the so-called Nature of Science (Guisasola & Morentin, 2007; Martínez-Chico, Jiménez-Liso & López-Gay, 2015; Mellado, 1996), on some educational approaches in teacher education (Criado & García-Carmona, 2011; García Barros, Martínez-Losada, Vega & Mondelo, 2000; Rivero, Hamed, Martín del Pozo, Solís, et al., 2013), or on the effects of teachers' self-assessment of their personal experience in pre-service education (Cantón, Cañón & Arías, 2013; García-Barros, Bugallo & Fuentes, 2013). Up to our knowledge there are no recent studies to assess teachers' elementary science knowledge, although this has been done with their

Received: 03-08-2017 Accepted: 07-09-2018

mathematical knowledge (Gómez & Gutiérrez-Gutiérrez, 2014; Sáenz-Castro, 2007).

Due to the lack of recent data, it seems interesting to conduct an empirical study aimed at evaluating Spanish pre-service elementary teachers' education in basic science, as it is done at the university. In-service assessment will be done later.

Goals

Therefore, the main goal of the present study is to evaluate the conceptual and procedural levels of basic science in pre-service teachers, before and after receiving the science education included in the university degrees. We focused our study on the Valencian Community as a relevant example of Spanish political and cultural lands, with a great population in Spanish terms. We aimed at achieving high external reliability in this study, providing suggestions to make suitable changes in present pre-service teachers' science education programmes.

Related to this main goal is the intention of obtaining detailed information about the level of efficacy of the nowadays science education offered in the Valencian public university degrees in teacher education. Obtaining some suggestions on how to significantly increase the lowest initial levels of science domain along the degree will be good. This implies taking into account some factors which could influence the initial and final levels of science domain. In the present study we considered the prior degree with which the student enters to the pre-service teacher training university degree (baccalaureate, vocational or professional training, another university degree), and the academic branch or specialization in his/him pre-university education (science & technology vs social sciences & humanities). Glaser (1995; see also Bransford, Brown & Cocking, 1999) pointed on some relevant findings in educational research on how improve the design and structure of learning environments. One of the particular findings is that students' previous knowledge and intuition are an important source of cognitive skills, which may conduce to new learning. Therefore, the student prior knowledge should be taken into account.

Some previous research found poor results and hardly dependent on the students' pre-university degree (Pérez de Landazábal, Benegas,

Received: 03-08-2017 Accepted: 07-09-2018

Cabrera, et al., 2010). However, this research involved students in the experimental science branch only. In the present study the participants show heterogeneity in their degrees as well as in their academic branches in the pre-university education.

In addition to the previous education, the total amount of classroom hours offered in the degree programmes could also be a factor influencing the final levels of science domain achieved. In Spain there is some degree of freedom for the universities to determine the total amount of classroom hours in different subjects, in particular those of science and science education for teachers. In that way different universities can adjust their programmes to particular goals or needs. More classroom hours devoted to science in the degree, would cause a significant improvement in pre-service teachers' science knowledge? Having reliable and current data on the answer could help authorities at the universities to make difficult decisions, as the increasing the hours of one subject diminishes the hours of others.

The assessment of the basic science knowledge

The present work takes the Miller's proposal of three dimensions for SL (1989): (1) Understanding of the processes and methods Science uses to test its models to explain reality; (2) basic vocabulary to properly name science and technology concepts; (2) some comprehension of the impact Science has or can have on nowadays society. The two first dimensions refer to the knowledge about Science (understanding and applications) and the third one implies the use of the science knowledge in social scenarios in a responsible way.

Some instruments have been designed based on these previous dimensions. For instance, Lauksch and Spargo (1996) developed the Basic Scientific Literacy Test made of 472 true/false items, covering 240 important ideas in science and attitudes towards science. This test is based on the Science for all Americans report (AAAS, 1989). Other instruments based on these dimensions have been used in different studies promoted and conducted by high quality institutions, as the National Science Foundation (NSF) in the US. In a long way international study on adult citizens' basic scientific culture (i.e. SL), a questionnaire made of 9 items was used to assess the factual knowledge in sciences (NSB, 2014). The

Received: 03-08-2017 Accepted: 07-09-2018

level of understanding of science processes was assessed by means of two questions on probability, and another additional open question on the meaning of an experimental scientific work. Similar questionnaires were used in the Eurobarometers (Special Eurobarometer 43, 76, 224) for assessing the factual science knowledge in Europe. In these studies, the level of comprehension of science processes were assessed with three questions about the experimental method, the role and use of control groups, and the concept of probability. In Spain, the Fundación para la Ciencia y la Tecnología (the Science and Technology Foundation; FECYT) took the european model as a reference to conduct their studies. After some changes made after initial studies, in 2014 the declarative science knowledge was evaluated using true/false items, and the science processes were evaluated with a single item on the role of a control group in a clinical study (FECYT, 2015).

Brossard and Shanahan (2006) focused on the second Miller's dimension, the understanding of science and technology terms and concepts, to elaborate their instrument. A total of 31 fill-in-the-blank items were validated. In each item, the definition of a particular concept is given, and the blank has to be filled-in with the correct term for this concept. The 31 concepts were extracted from the most frequent ones appearing in Mass Media. The authors compared the results obtained using this instrument to the ones obtained using the NSF questionnaire to assess SL. When age, gender and level of science education were controlled, the resulting correlation was considered high enough. In a similar vein, Rundgren, Rundgren, Tseng, Lin and Chang (2010) designed the instrument SLiM to assess SL. This questionnaire is made of 50 multiple option items. The most frequent science concepts in textbooks and newspapers in Taiwan were selected. SLiM showed good reliability with Taiwanese students from 7th to 10th grades.

The above studies and the instruments they used inspired the present work conducted in the Spanish context and focalised on pre-service Primary teachers (and not on the general adult population). Up to our knowledge, there is not any high reliability study of this type in recent times.

Received: 03-08-2017 Accepted: 07-09-2018

Method

Design

To achieve our goals, the basic conceptual knowledge was split into two components: (a) knowledge about the science content, of conceptual nature; (b) procedural knowledge, associated to the mental processes implied in building and validating Science. We defined the following dependent variables:

1) Science Conceptual knowledge at the level of primary school
2) Science Basic Procedural knowledge

We tried the two components not to interfere each other, so the knowledge about science knowledge was not a pre-requisite to procedural knowledge, or the other way round. Therefore, the instruments to assess both components were carefully selected (see below).

A factorial synchronic *ex-post facto* design was conducted, with a between-subjects factor, the Phase of the study with two levels: before/after receiving the compulsory education in science subjects in the university degree in pre-service teacher training.

Complementary, we also considered the possible effects related to other factors as the specific "conceptual blocks" or the "processual categories" which can be differentiated in the science knowledge, as within-subject factors. The Spanish educational legislation mentioned 4 conceptual blocks to be addressed in primary science: I) the Environment and its Preservation[1]; II) the Diversity of Life; III) Health and personal development; IV) Matter and Energy. Related to the basic processual knowledge, we considered 5 different categories: I) Identifying and Controlling Variables; II) Formulating Hypotheses; III) Operational definitions; IV) Interpreting Graphs and Data, V) Experimental Design.

In order to achieve our specific goals, we considered other between-subject factors: "prior degree used to access the pre-service teacher

(1) The science content included in "the Environment and its Preservation" was first included in the academic matter "the Knowledge about the Environment", but nowadays it is included in the content block " the World where we all Live" in the matter "Social Sciences", according to the legal Norm: Decreto 108/2014 (DOGV 7311/07.07.2014) for the Primary Education curriculum in the Valencian Community. As it is a content of social and experimental nature as well, we decided to maintain the block in the present study.

Received: 03-08-2017 Accepted: 07-09-2018

training degree": Baccalaureate & other University degrees versus Other degrees); "prior academic branch of specialization": Science & Technology versus Other specialities; "Compulsory curricular load in sciences" in the pre-service teacher training degree.

The three public universities in the Spanish Valencian Community offering the pre-service teacher training degree are: the University of Valencia, the University of Alicante and the University "Jaume I" in Castellón. However, at the moment of this study, the University of Valencia offered 180 classroom hours in science subjects versus the 120 classroom hours offered in universities of Alicante and Castellon. In addition, the distribution of these classroom hours was also different. At the University of Jaume I students had to pursue two compulsory science matters: Physics and Chemistry Education (60 hours; 2nd year) and Natural Sciences Education (60 hours; 2nd year). The University of Alicante also offered two compulsory science matters to their students: Teaching and learning Science I and II (both having 60 hours; in the first and the second half of the 2nd year). As the University of Valencia concerns, students had to pursue three different matters: Science for Teachers (90 hours; second year); Science Education I: Matter, Energy and Machines (45 hours, third year); Science Education II: Environment, Biodiversity and Health (45 hours; four year). In this last university, Science for Teachers was devoted to basic science content whereas the two Science Education matters focused on developing teaching skills. This division between disciplinary and teaching education in science did not appear in the other two universities, Alicante or Jaume I (Castellón).

Sample

The populations considered were composed by all students in the pre-service Primary teacher training at the public universities in the Valencian Community (Spain). In particular, the populations in the present study were the subsets of students in the grades whereby the matters related to science initiate (Phase I: previous studying the first compulsory science matter at the degree) and ends (Phase II: just after the last compulsory science matter was pursued). The population sizes were computed from

Received: 03-08-2017 Accepted: 07-09-2018

the data given by the different universities in their official webs[2]. As a result, a total amount of 1210 male and female students made the target population in Phase 1. As the three public universities informed that their dropout rates were about 10 percent, the population in Phase 2 was estimated to be made of 1100 male and female students.

Sampling was made in the different pre-service teacher training faculties at the three public universities in the Valencian Community.

Table I shows the resulting samples associated to the two measures obtained in the present study (basic science conceptual knowledge / procedural knowledge) as well as the sampling error at the 95% confidence level.

TABLE I. Population, simple and sampling error in Phases I and II.

Phase of the study	Population (N)	Concepts		Procedures	
		Sample	Sampling Error	Sample	Sampling Error
1 (before)	1210	289	5,0%	308	5,2%
2 (after)	1100	270	4,8%	286	5,0%

Source: authors' elaboration

The classroom group, and not the single student, was considered to be the sampling unit. At each university and in each involved grade the participant groups were selected and individual and voluntary information was requested in ordinary classroom sessions of science subjects.

The classroom groups of informants were not strictly selected at random, but the procedure contained the random factor associated to the day/hour chose to obtain the data at each university.

Looking at the days and hours the science lessons were given at each university, we select one particular day/hour at random to go and obtain the data. Thus, although we did not follow a strictly random procedure to choose the classroom groups, there were an at random factor in the

(2) University of Alicante: http://cvnet.cpd.ua.es/webcvnet/planestudio/planestudiond.aspx?plan=C254#tramites
University Jaime I : http://ujiapps.uji.es/estudis/oferta/base/graus/2015/mestre-primaria/accedir/oferta-procediment-admissio/
University of Valencia: http://www.uv.es/uvweb/universitat/ca/estudis-grau/grau-1285846094474/Titulacio.html?id=1285847460730

Received: 03-08-2017 Accepted: 07-09-2018

selection, as each different groups of students received their science lessons in different days/hours.

Percentages of informants present in the selected classrooms (i.e present / total students in the group) were: 88 percent in Alicante, 91 percent in Castellon, and 97 percent in Valencia. The sampling ratio of the classroom groups was close of 50 percent in both Phases. However, due to the time the teachers gave us and the informants' fatigue (both instruments have 30 items with 4 options), every participant was only assessed in concept knowledge or in procedural knowledge. The instruments to assess concept knowledge and to assess procedural knowledge were both randomly distributed among the students in every classroom.

The sample distributions are shown in Table II.

TABLE II. Sample distributions according to the factors considered.

	Phase 1 (before)				**Phase 2 (after)**			
	Prior Specialization				**Prior Specialization**			
Accesing Degree	**Sci&Tech**		**Other**		**Sci&Tech**		**Other**	
Bacc & Univ	M:10	W:33	M:57	W:147	M:12	W:36	M:36	W:137
Prof. Training & Other	M: 0	W: 2	M: 7	W: 27	M: 1	W: 4	M:10	W: 30

Source: authors' elaboration
M: men; W: women. Experimental mortality N=10

Instruments

The instrument used to obtain the participants' basic science conceptual knowledge was developed and validated by Verdugo-Perona, Solaz-Portolés and Sanjosé (2016) for Spanish pre-service Primary teachers. A shortening, improvement and adaptation process was followed. First, items having extreme scores (more than 0,85 or less than 0,15 out of 1), or those having low discrimination indexes (less than 0,2) were taken off. The resulting 30 items questionnaire was reliable enough given the great variability in the concepts included (Kurder-Richardson's 20th formula; KR= ,7).

Received: 03-08-2017 Accepted: 07-09-2018

Three important features of this instrument are: a) the conceptual blocks considered in the Spanish curriculum for Primary Education are well represented; b) answering any of the items only requires having conceptual knowledge but not another type of knowledge; c) Only one concept is involved in each item. The second feature implied that memory is necessary to give the correct answer, but the 4 answer options minimize the influence of student's memory.

A 0/1 scoring is used for wrong/ correct answer in each item. Wrong options are not penalized and then, the total scoring is obtained by adding the individual item scores (max= 30 points).

Items in the questionnaire belong to the four content blocks and were elaborated after a deep study of the official normative documents and the analysis of the most popular and best sold textbooks (see the Design sub-section). The Pearson' correlations between pairs of content blocks were all significant (,25< r < ,34; p< ,001).

The students' procedural knowledge was obtained using a different instrument elaborated and validated by Monde-Monica (2005) and of ease administration. This questionnaire is made of 30 multiple option items and it has two important features: a) the questionnaire is not biased towards any particular science (physics, biology, etc.); b) the correct answer for any item can be given without any specific science conceptual knowledge. The global score is also obtained by adding the individual item scores (0/1). Each item is associated to one of the 5 categories considered for the procedural knowledge (see the Design sub-section). Despite the "Experimental Design" category, the other categories significantly correlated (,28 <r < ,30; p<,001). "Experimental Design" only weekly correlated with "Operational definitions" (r= ,10; p= ,028).

Global reliability of this instrument was considered good enough as well (KR-20= ,7).

In both instruments used the individual scores for each conceptual block (in the questionnaire assessing the science conceptual knowledge) or for each procedural category (in the questionnaire assessing the science procedural knowledge) did not reach good reliability (KR-20< ,6 in any case). This was probably due to the low amount of items involved in each block (6-9 items) or category (3-8 items). Hence the scores of single conceptual blocks or single procedural categories can only suggest trends about the students' knowledge in each one. Conclusions brought out from these scores will have limited reliability and they will need

Received: 03-08-2017 Accepted: 07-09-2018

further confirmation. However, the outcomes and analyses made from global scores of both instruments will have enough reliability.

In some analyses and in order to compare blocks or categories with different number of items, the scores were normalized to the 0-1 rank.

The Appendix shows some examples of items in both instruments.

Procedure

After permissions obtained from the implied Education schools in each university, one of the authors (JJVP) went to every selected intact group and ask students voluntary participation. The same day the data about conceptual and procedural knowledge were obtained before and after the compulsory science education subjects were pursued. No other data were obtained in intermediate moments or phases in the science education process at the university degree.

Data collection took place in the first or second hour in the normal schedule of the classroom sessions in the morning or the evening shift. At the beginning students were informed about the main academic goals of the present study and their voluntary participation was requested. The typical time for completion in the case of the science concepts questionnaire was 40 min, and 50 min in the case of the science processes questionnaire.

Data collected were organized using Excel, and were processed for statistical analyses with SPSS v.22. ANOVA was used to evaluate the effects caused by single or joined factors on the dependent variables (the scores). As a complementary analysis, the paired-t test (Student's test) was also used for Phase1/ Phase2 comparisons.

The confidence level for rejecting the null hypothesis was p= ,05 in any analysis.

Results

Table III gives the main mean values and standard deviations (in brackets) of the scores obtained for the conceptual and also for the procedural science knowledge (both maximum values= 1).

Received: 03-08-2017 Accepted: 07-09-2018

TABLE III. Mean values (SD) for basic science conceptual and procedural knowledge, according to the factors considered.

Factors	Conceptual knowledge		Procedural knowledge	
	Phase 1	Phase 2	Phase 1	Phase 2
Global	0,50 (0,14)	0,60 (0,12)	0,65 (0,13)	0,68 (0,12)
Block/Category				
I	0,50 (0,21)	0,63 (0,22)	0,76 (0,21)	0,82 (0,19)
II	0,54 (0,20)	0,62 (0,17)	0,52 (0,21)	0,54 (0,19)
III	0,45 (0,20)	0,55 (0,18)	0,59 (0,23)	0,61 (0,20)
IV	0,53 (0,23)	0,60 (0,21)	0,78 (0,18)	0,82 (0,17)
V	----	----	0,43 (0,26)	0,44 (0,25)
Prior Speciality				
Experimental Science	0,58 (0,14)	0,67 (0,11)	0,69 (0,13)	0,72 (0,10)
Other fields	0,49 (0,14)	0,58 (0,12)	0,65 (0,13)	0,67 (0,13)
Prior degree				
Baccalaureate & other University degrees	0,50 (0,14)	0,60 (0,13)	0,65 (0,13)	0,69 (0,12)
Other Degrees	0,49 (0,14)	0,58 (0,12)	0,65 (0,14)	0,65 (0,13)
Curricular load				
Univ. Valencia	0,53 (0,15)	0,64 (0,11)	0,66 (0,13)	0,71 (0,10)
Univ. Castellón & Alicante	0,48 (0,14)	0,57 (0,13)	0,65 (0,13)	0,67 (0,13)

Source: authors' elaboration

The evolution of the basic science conceptual knowledge

In Table III the global score for conceptual knowledge increasing from 0,50 in Phase 1 (before the science education in the degree) to 0,60 in Phase 2 (after the science education in the degree) of the study. This implies a relative gain (out the maximum possible) of:

$$\frac{(0{,}60 - 0{,}50)}{(1 - 0{,}50)} = 0{,}20$$

If the maximum possible (ideal) improvement was 0,50, the obtained improvement was only the 20 percent of this quantity (0,10 is 20 percent of 0,50).

Received: 03-08-2017 Accepted: 07-09-2018

ANOVA for the global score taking the Phase (1/2) of the study as the within-subjects factor, showed this increasing was significant with a medium-to-high effect size ($F(1,557)= 71,388$; $p< ,001$; $\eta^2= ,114$). This is an additional meaning to the 20 percent of relative gain.

As a complementary information, the Phase X Block was not significant (Wilks' Lambda: $F(3,555)= 1,958$; $p> ,10$). Keeping in mind that the average score for each of the concept blocks was not reliable enough, this result suggested that the increases in the conceptual knowledge of the different concept blocks were statistically similar.

The evolution of the basic science procedural knowledge

The different types of procedural skills considered also showed changes in their mean scores from Phase 1 to Phase 2. The global change was only of 0,03 points (from 0,65 to 0,68) meaning a relative gain of:

$$\frac{(0,68-0,65)}{(1-0,65)} = 0,086$$

Therefore, the relative gain was near 9% of the maximum, ideal gain (0,35 points).

ANOVA computed for the global average score with the Phase (1/2) as the within-subjects factor showed that this change was also significant due to the systematic gains in the sample, although the effect size was very small ($F(1,592)= 9,0091$; $p= ,03$; $\eta^2= 0,015$).

A complementary multivariate analysis with the Phase and the procedural Category as the factors showed a non-significant Phase X Category interaction, (Wilks' Lambda: $F(4,589)= 2,021$; $p> ,05$). With the caution due to the lack of reliability of the average scores for each procedural category, this result suggested that the changes observed in the different procedural categories were similar (and very small) in statistical terms.

Received: 03-08-2017 Accepted: 07-09-2018

Effects due to the academic Branch previous to the pre-service teacher training

Before receiving the science subjects at the degree (Phase 1), the science conceptual knowledge showed a significant dependence from the speciality or branch of the previous studies (Science& Technology/ Other), as ANOVA confirmed ($F(1,283)= 18,960$; $p<,001$; $\eta^2= 0,63$) with a medium-small effect size.

The same happened with the science processual knowledge before the science subjects pursued at the university degree ($F(1,306)= 4,343$; $p= ,038$; $h^2= 0,014$) although the effect size was clearly small.

When all the compulsory science subjects were carried out (Phase 2), the differences caused by the prior specialization in both types of science knowledge were still significant for the conceptual knowledge ($F(1,268)= 22,950$; $p<,001$; $\eta^2= 0,079$) as also for the procedural one ($F(1,284)= 9,073$; $p= ,003$; $\eta^2= 0,031$).

The Phase X Prior Speciality was not significant for both types of science knowledge ($F<1$). Hence, the observed "Science-Other" differences in Phase 1 were of similar magnitude than the observed in Phase 2, after all the compulsory science subjects were pursued in the pre-service teacher training at the university.

Effects due to the prior degree

The corresponding ANOVA for the average scores in science conceptual and procedural knowledge were computed taking the type of prior degree as a between-subjects factor. We considered Baccalaureate & other University degrees *versus* Other degrees. There were not significant differences between these two types of prior degrees in Phase 1 or in Phase 2 ($p> ,10$ in both cases). And neither was for the Phase X Prior Degree interaction ($p> ,10$). Globally speaking, the knowledge of students entering in the teacher training degree from Baccalaureate or other university degree was similar to the improvement of students entering from professional training degrees or other possibilities.

Received: 03-08-2017 Accepted: 07-09-2018

Effects due to the different amount of classroom hours for science subjects

As mentioned before, there are differences among the three considered universities in the total amount of classroom hours devoted to compulsory science subjects. In the university of Valencia, the curricular load in science subjects is greater than in the other two (Jaume I in Castellon, and University of Alicante). The distribution of these hours is different too. In both Phases of the study the average science knowledge scores were higher in the university of Valencia than in the other two universities together. However, the gain between Phase 1 and Phase 2 was of similar magnitude in statistical terms. In fact, the Phase X Curricular load in science (Valencia vs. Castellon & Alicante) was not significant neither for the conceptual knowledge nor for the procedural knowledge ($p > ,10$ in both cases).

Discussion and Conclusions

The goal of this study was to assess whether pre-service primary school teachers in the Valencian Community were scientifically literate enough to make literate to their pupils. For that purpose, this study considered the level of basic scientific knowledge that undergraduates in pre-service teacher education degree from the Valencian Community present before and after they coursed the compulsory science education subjects. Results in this Community do not have to be identical to those in other Spanish communities, but they may provide us an indication about the level of scientific literacy in Spanish pre-service primary teachers. In the present study the sampling error was about ±5 percent (at the 95 percent of confidence level), quite similar to the study conducted by FECYT (2015) for the Valencian Community (±5,03 percent). Two of the Miller's (1998) scientific literacy dimensions were taken into account to assess SL: knowledge in basic scientific concepts and the understanding of scientific processes. These two dimensions were assessed using validated instruments (Verdugo-Perona, Solaz-Portolés & Sanjosé, 2016; Monde-Monica, 2005).

Results showed that most future primary teachers accessed to the degree with poor conceptual knowledge, considering that the assessment was made only on basic science content commonly found in Primary

Received: 03-08-2017 Accepted: 07-09-2018

school textbooks (neither secondary school nor university). The score averages increased significantly after the science training received at University but effects size were medium in conceptual knowledge and very small in procedural knowledge.

This work does not let us know the reasons why this improvement in procedural knowledge is on small. Even so, one of the factors that it should be borne in mind for future studies could be the traditional lack of students' work based on inquiry learning throughout the pre-service training degree, as it has been found in other contexts as well (Vilches & Gil-Pérez, 2007; Abd-El-Khalick et al, 2004).

The conceptual knowledge showed by participants yielded low averages, considering that the evaluation was about basic science contents, which are typical in Primary School Syllabus. The average increased significantly after mandatory training in the degree programme, with a medium effect size.

In the current study, the levels of conceptual knowledge were similar to those obtained in previous general studies (Eurobarometer 224). However, they are considerably lower than the ones obtained in the study conducted by FECYT (2015). The differences may be due to the level of difficulty of the questions used in both studies. On one hand, FECYT uses items made up of complete ideas that participants must "recognise" as true or false. In our research participants must discriminate the correct option among other three distractors belonging to the same conceptual field. On the other hand, the questionnaire used in this article (Verdugo-Perona, Solaz-Portolés & Sanjosé, 2016) involves a larger number of concepts than the one used by FECYT. However, this greater difficulty is coherent with the target population of the instrument used: future primary teachers.

Trainee teachers also showed a higher level of science procedural domain compared to the one showed by general population in FECYT's study. This latest study is based on one only fundamental and general item ("To understand the importance of comparing experimental groups with control groups"). Unlike this study, and according to the more specific features of our participants (university students, trainee teachers), the present study used a validated questionnaire covering a variety of situations (Monde-Monica, 2005), which allowed a deep and reliable assessment.

Received: 03-08-2017 Accepted: 07-09-2018

The analysis of the results by content blocks or by procedural categories enabled the identification of those with poorer outcomes, thus requiring special attention. As regards the science procedures, category V ("Experimental Design") obtained the lower scores, and the students' competence in this category hardly improved in the degree. This result could be explained, first because of the difficulty in acquiring competences of this type, inherent to researchers and not to teachers; and second, perhaps because during the degree programme they did not have many opportunities of working on experimental designs, that it is a time-consuming task. Despite the fact that the use of inquiry-based methodologies for teaching science has been proposed for decades (Osborne y Dillon, 2008), the competences needed for this are not usually addressed in teacher training so far (Tierno, Gavidia y Tuzón, 2018; Haefner & Zembal-Saul, 2004). Of course, we should be aware of the urgent problems university teachers deal with, as their students' poor conceptual knowledge for instance, that strongly limits students' teaching education in science methods to shallow approaches. In the present study participants showed a basic conceptual knowledge that was clearly improvable, and an especially low score in block III ("Health and personal development"), traditionally receiving a large amount of resources that, do not produce the expected results.

Additional analyses were made involving some factors that could clarify the global results obtained. The type of prior degree used to access to the pre-service teacher training degree (baccalaureate and other university graduates, versus other options as vocational training degrees, special access for people over 25, 40, 45 y.o., etc.) did not significantly influence the results obtained in conceptual knowledge or in procedural knowledge. This suggests that students who access to the university from vocational training or access as mature students show a level of knowledge similar to students entering with a baccalaureate.

However, the prior speciality or academic branch caused significant differences in the scores. Indeed, students accessing from science & technology prior degrees initiated the pre-service teacher training degree with a significantly better knowledge than the other students, both in science concepts and science processes. These students took more profit from the science education received in the teacher training programme and achieved higher levels of science knowledge than the other students at the end of the training. This beneficial impact of the previous knowledge

Received: 03-08-2017 Accepted: 07-09-2018

is well known and has been replicated in multiple educational studies (Braasch & Goldman, 2010; Irrazabal, 2010). This means that particular attention should be put on students entering the teacher training degree with the lower levels of science knowledge, trying to help them improve their motivation, effort, attention and enjoyment of science.

Finally, in the university of Valencia, the one with the highest number of classroom hours devoted to compulsory science subjects, students showed higher conceptual and procedural domain than students in the other two public universities considered here. This happened before and also after the science subjects were pursued in the different university degree programmes. However, the greatest amount of hours did not cause a greater gain in knowledge. This was especially true when comparing the universities of Valencia and Alicante. Obviously, the amount of classroom hours offered in the degree do not seem enough to explain students' learning. Teachers' instructional methodology, the way they manage their resources and the more or the less emphasis they place on some contents (conceptual, procedural or educational contents) are probably key factors as well, and they have not been addressed in the present evaluative study.

A factor that could contribute to explain the greater instructional efficiency of the university of Alicante compared to the university of Valencia is of instructional nature: in the university of Alicante students learn disciplinary science knowledge together with pedagogical science knowledge. Instead, students in Valencia course science content and science teaching in different matters. Placing the students in the role of teachers might cause a metacognitive effect, improving disciplinary science knowledge. This phenomenon has been raised in many studies (Dignath & Büttner, 2008).

Limitations of the study

Although it has enough external validity and reliability in terms of general results, this study includes some features defining its limitations.

First, the two considered components of scientific knowledge, conceptual and procedural knowledge, were independently assessed. While this may be technically important to conclude about those components separately, there is no doubt that scientific knowledge and

Received: 03-08-2017 Accepted: 07-09-2018

its daily life application are based on concepts and processes together, necessary to describe, understand or predict natural phenomena. The fact that the instrument for assessing the conceptual knowledge involved the student's memory could have some influence the low average obtained, even at the end of the training programme. Similarly, the instrument used to assess the knowledge of science processes does not involve specific conceptual knowledge, so the results obtained could be different if scientific concepts were included in the situations presented to the participants in the items.

Second, the instruments used here only ensure enough reliability in the global scores, but not in the particular scores of each scientific concept block or procedural category. Thus, as explained before, the results assessing each block or category should be considered as mere guidelines. Reliability could be only increased after sufficient replications or using more specific instruments for specific conceptual blocks or procedural categories (e.g. with largest number of well-cohered items).

Third, the assessment of the students' knowledge about the nature of science is still pending in this same population. This type of knowledge has received considerable attention (Fernández, Gil, Carrascosa & Cachapuz, 2002; Irez, 2006; Guisasola & Morentin, 2007; Acevedo-Díaz, 2008). It would be also interesting to assess pre-service teachers' pedagogical content knowledge in science (Blanco, Mellado & Ruiz, 1995; Loughran, Mulhall & Berry, 2008; Nilsson, 2008). Both types of knowledge are important to the future development of the citizens' scientific literacy.

The last remarkable limitation is the choice of a synchronous study, instead of a diachronic cross-sectional one, which would diminish the variance error by using the same informants throughout the research.

References

Abd-El-Khalick, F., Boujaoude, S., Duschl, R., Lederman, N. G., Mamlok-Naaman, R., Hofstein, A. & Tuan, H. L. (2004). Inquiry in Science Education: International perspectives. *Science Education*, *88*(3), 397-419. DOI:10.1002/sce.10118

Received: 03-08-2017 Accepted: 07-09-2018

Acevedo, J.A. (2004). Reflexiones sobre las finalidades de la enseñanza de las ciencias: Educación científica para la ciudadanía. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 1, 3-16.

Acevedo-Díaz, J. A. (2008). El estado actual de la naturaleza de la ciencia en la didáctica de las ciencias. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, *5*(2), 133-169.

American Association for the Advancement of Science [AAAS] (1989). *Science for all Americans.* Retrieved from: http://www.project2061.org/publications/sfaa/online/sfaatoc.htm].

Appleton, K. (2003). How do beginning primary school teachers cope with science? Toward an understanding of science teaching practice. *Research in science education*, *33*(1), 1-25.

Blanco, L. J., Mellado, V. & Ruiz, C. (1995). Conocimiento didáctico del contenido en Ciencias Experimentales y Matemáticas y Formación de Profesores. *Revista de Educación*, 307, 427-446.

Braasch, J. L. & Goldman, S. R. (2010). The role of prior knowledge in learning from analogies in science texts. *Discourse Processes*, *47*(6), 447-479. DOI:10.1080/01638530903420960

Bransford, J., Brown, A. & Cocking, R. (Eds.). (1999). How people learn. Washington, DC: National Academy Press.

Brossard, D. & Shanahan, J. (2006). Do they know what they read? *Science Communication*, *28*(1), 47-63. DOI:10.1177/1075547006291345

Dignath, C. & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, *3*(3), 231-264. DOI:10.1007/s11409-008-9029-x

Cantón, I., Cañón, R. & Arías, A. R. (2013). La formación universitaria de los maestros de Educación Primaria. *Revista Interuniversitaria de Formación del Profesorado*, *27*(1), 45-63.

Cañal, P. (2000). El conocimiento profesional sobre las ciencias y la alfabetización científica en Pri- maria. *Alambique Didáctica de las Ciencias Experimentales*, *24*, 46-56.

Cañal, P. (2008). ¿Cómo orientar la formación inicial del profesorado de primaria en didáctica de las ciencias experimentales? En M. R. Jiménez Liso (Ed.). *Ciencias para el mundo contemporáneo y formación del profesorado en Didáctica de las Ciencias Experimentales* (pp. 256-263). Almería: Ed. Univ. Almería.

Received: 03-08-2017 Accepted: 07-09-2018

Criado, A. M. & García-Carmona, A. (2011). Las experiencias prácticas para el conocimiento del medio (natural y tecnológico) en la formación inicial de maestros. *Investigación en la Escuela*, 74, 73-88.

Decreto 108/2014, de 4 de julio, del Consell, por el que establece el currículo y desarrolla la ordenación general de la educación primaria en la Comunitat Valenciana. [2014/6347]. Publicado en DOGCV, num 7311/07 07 2014, p 16510. Retrieved from: http://www.docv.gva.es/datos/2014/07/07/pdf/2014_6347.pdf.

FECYT (2015). Percepción Social de la Ciencia y la Tecnología en España. Madrid. Retrieved from: http://www.fecyt.es/es/publicacion/percepcion-social-de-la-ciencia-y-la-tecnologia-2014.

Fernández, I., Gil, D., Carrascosa, J. & Cachapuz, A. (2002). Visiones deformadas de la ciencia transmitida por la enseñanza. *Enseñanza de las ciencias*, *20*(3), 477-488.

García Barros, S., Bugallo, Á. & Fuentes, M. J. (2013). Los objetivos de las ciencias en primaria y las necesidades formativas, vistas por los maestros en formación. *Enseñanza de las Ciencias*, special number, 1460-1466.

García Barros, S., Martínez Losada, C., Vega, P. & Mondelo, M. (2000). Propuesta de intervención para la formación inicial del profesorado de Educación Primaria en Ciencias Experimentales. *Revista Interuniversitaria de Formación del Profesorado*, 38, 153-165.

García-Ruiz, M. & Sánchez, B. (2006). Las actitudes relacionadas con las Ciencias Naturales y sus repercusiones en la práctica docente de profesores de primaria. *Perfiles Educativos*, 114, 61-89.

Gil, D. & Vilches, A. (2006). Educación ciudadana y alfabetización científica: mitos y realidades. *Revista Iberoamericana de Educación*, 42, 31-53.

Glaser, R. (1995). Application and theory: Learning theory and the design of learning environments. Paper presented at the 23rd International Congress of Applied Psychology, July 17-22, 1994, Madrid, Spain.

Gómez, P. & Gutiérrez-Gutiérrez, A. (2014). Conocimiento matemático y conocimiento didáctico del futuro profesor español de primaria. Resultados del estudio TEDS-M. In M. T. González, M. Codes, D. Arnau & T. Ortega (Eds.), *Investigación en Educación Matemática XVIII* (pp. 99-114). Salamanca: SEIEM.

Received: 03-08-2017 Accepted: 07-09-2018

Guisasola, J. & Morentin, M. (2007). ¿Comprenden la naturaleza de la ciencia los futuros maestros y maestras de Educación Primaria? *Revista Electrónica de Enseñanza de las Ciencias*, *6*(2), 246-262.

Haefner, L. A. & Zembal-Saul, C. (2004). Learning by doing? Prospective elementary teachers' developing understandings of scientific inquiry and science teaching and learning. *International Journal of Science Education, 26*(13), 1653-1674. DOI:10.1080/0950069042000230709

Harlen, W. & Holroyd, C. (1997). Primary teachers' understanding of concepts of science: impact on confidence and teaching. *International Journal of Science Education*, *19*(1), 93-105. DOI:10.1080/0950069970190107

Hurd, P. D. (1998). Scientific Literacy: New minds for a changing world. *Science Education*, *82*(3), 407-416. DOI:10.1002/(SICI)1098-237X(199806)82:3<407::AID-SCE6>3.0.CO;2-G

Irez, S. (2006). Are we prepared? An assessment of preservice science teacher educators' beliefs about nature of science. *Science Education*, *90*(6), 1113-1143. DOI:10.1002/sce.20156

Irrazabal, N. (2010). La comprensión de textos expositivos en estudiantes universitarios: la función del conocimiento previo. *Revista de Psicología*, *12*, 7-21.

Jarvis, T. & Pell, A. (2004). Primary teachers' changing attitudes and cognition during a two-year science in-service programme and their effect on pupils. *International Journal of Science Education*, *26* (14), 1787-1811. DOI:10.1080/0950069042000243763

Juriševič, M., Glažar, S. A., Pučko, C. R. & Devetak, I. (2008). Intrinsic Motivation of Pre-service Primary School Teachers for Learning Chemistry in Relation to their Academic Achievement. *International Journal of Science Education*, *30*(1), 87-107.

Kang, NH. (2007). Elementary teachers' teaching for conceptual understanding: learning from action research. *Journal of Science Teacher Education*, *18*(4), 469-495. DOI: 10.1007/s10972-007-9050-y.

Krugar, C., Summers, M. & Palacio, D. (1990) A survey of primary school teachers' conceptions of force and motion. *Educational Research*, *32*(2), 83-95. DOI:10.1080/0013188900320201

Laugksch, R.C. & Spargo, P. (1996). Construction of a paper-and- pencil test of basic scientific literacy goals recommended by the American Association for the Advancement of Science. *Public Understanding of Science*, *5*(4), 331–359. DOI:10.1088/0963-6625/5/4/003

Received: 03-08-2017 Accepted: 07-09-2018

Lloyd, J. K., Smith, R. G., Fay, C. L., Khang, G. N., Wah, L. L. K. & Sai, C. L. (1998). Subject knowledge for science teaching at primary level: a comparison of pre-service teachers in England and Singapore. *International Journal of Science Education*, *20*(5), 521-532. DOI:10.1080/0950069980200502

Loughran, J., Mulhall, P., & Berry, A. (2008). Exploring pedagogical content knowledge in science teacher education. *International Journal of Science Education*, *30*(10), 1301-1320. DOI:10.1080/09500690802187009

Martínez-Chico, M., Jiménez Liso, M. R. & López-Gay, R. (2015). Efecto de un programa formativo para enseñar ciencias por indagación basada en modelos, en las concepciones didácticas de los futuros maestros. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias, 12*(1), 149-166. DOI:10.25267/Rev_Eureka_ensen_divulg_cienc.2015.v12.i1.10

Mellado Jiménez, V. (1996). Concepciones y prácticas de aula de profesores de ciencias, en formación inicial de primaria y secundaria. *Enseñanza de las Ciencias*, *14*(3), 289-302.

Miller, J. D. (1989). Scientific literacy. Paper presented at the annual meeting of the American Association for the Advancement of Science, San Francisco. January 1989.

Miller, J. D. (1998). The measurement of civic scientific literacy. *Public understanding of science*, *7*(3), 203-223. DOI:10.1088/0963-6625/7/3/001

Monde-Monica, K.M. (2005). Development and validation of a test of integrated science process skills for the further education and training learners. Master Thesis. University of Pretoria, South Africa. Retrieved from: http://repository.up.ac.za/bitstream/handle/2263/24239/dissertation.pdf?sequence=1

Murphy, C. & Smith, G. (2012). The impact of a curriculum course on pre-service primary teachers' science content knowledge and attitudes towards teaching science. *Irish Educational Studies*, *31*(1), 77-95. DO I:10.1080/03323315.2011.634061

National Research Council. (1996). National Science Education Standards. Washington, DC: National Academy Press. Retrieved from: http://www.nap.edu/catalog/4962/national-science-education-standards.

National Science Board (2014). Science and Technology: Public attitudes and understanding. En *Science and Engineering Indicators 2014* (7.20-7.23). Arlington VA: National Science Foundation (NSB-14-01).

Received: 03-08-2017 Accepted: 07-09-2018

Retrieved from: http://nsf.gov/statistics/seind14/index.cfm/chapter-7/c7h.htm.

Nilsson, P. (2008). Teaching for Understanding: The complex nature of pedagogical content knowledge in preservice education. *International Journal of Science Education*, *30*(10), 1281-1299. DOI:10.1080/09500690802186993

Osborne, J. & Dillon, J. (2008). *Science education in Europe: Critical reflections* (Vol. 13). London: The Nuffield Foundation

Papageorgiou, G., Stamovlasis, D. & Johnson, P. M. (2010). Primary Teachers' Particle Ideas and Explanations of Physical Phenomena: Effect of an in-service training course. *International Journal of Science Education*, *32*(5), 629-652. DOI:10.1080/09500690902738016

Pérez de Landazábal, M.C., Benegas, J., Cabrera, J.S., Espejo, R., Macías, A., Otero, J., Seballos, S. & Zavala, G. (2010). Comprensión de conceptos básicos de la Física por alumnos que acceden a la universidad en España e Iberoamérica: limitaciones y propuestas de mejora *Latin American Journal of Physics Education, 4* (3), 655-668. Retrieved from: http://www.lajpe.org.

Porlán, R.; Martín del Pozo, R.; Rivero, A.; Harres, J.; Azcárate, P. & Pizzato, M. (2010). El cambio del profesorado de ciencias I: Marco teórico y formativo. *Enseñanza de las Ciencias*, *28*(1), 31-46.

Rivero, A., Hamed, S., Martín del Pozo R., Solís, E. Fernández, J., Porlán R., Rodríguez, F., Solís, C., Azcárate, P. & Ezquerra, Á. (2013). La formación inicial de maestros de primaria: qué hacer y cómo en didáctica de las ciencias. *Enseñanza de las Ciencias*, *número extra*, 3045-3050.

Rundgren, C.-J., Rundgren, S.-N. C., Tseng, Y.-H., Lin, P.-L. & Chang, C.-Y. (2010). Are you SLiM? Developing an instrument for civic scientific literacy measurement (SliM) based on media coverage. *Public Understanding of Science, 20*(10), 1–15. DOI:10.1177/0963662510377562. Retrieved from: https://www.researchgate.net/profile/Chun_Yen_Chang/publication/247156553_Are_you_SLiM_Developing_an_instrument_for_civic_scientific_literacy_measurement_SLiM_based_on_media_coverage/links/545972bd0cf2bccc4912bcee/Are-you-SLiM-Developing-an-instrument-for-civic-scientific-literacy-measurement-SLiM-based-on-media-coverage.pdf

Sáenz Castro, C. (2007). La competencia matemática (en el sentido de PISA) de los futuros maestros. *Enseñanza de las Ciencias*, *25*(3), 355-366.

Special Eurobarometer 43 (1990). Les Europeens, la Science et la Technologie. Commission of the European Communities. Paris. Retrieved from: http://ec.europa.eu/COMMFrontOffice/PublicOpinion/index.cfm/Survey/getSurveyDetail/yearFrom/1988/yearTo/1990/surveyKy/101].

Special Eurobarometer 224 (2005). Europeans, Science and Technology. Commission of the European Communities. Brussels. Retrieved from: http://ec.europa.eu/COMMFrontOffice/PublicOpinion/index.cfm/Survey/getSurveyDetail/yearFrom/2004/yearTo/2007/surveyKy/447/p/3].

Special Eurobarometer 76 (1993). Europeans, Science and Technology. Commission of the European Communities. Brussels. Retrieved from: http://ec.europa.eu/COMMFrontOffice/PublicOpinion/index.cfm/Survey/getSurveyDetail/yearFrom/1991/yearTo/1993/surveyKy/131]

Summers, M. & Krugar, C. (1992). Research into English primary teachers' understanding of the concept of energy. *Evaluation and Research in Education*, 6, 95-111. DOI:10.1080/09500799209533321

Tierno, S., Gavidia, V. & Tuzón, P. (2018). Radiografía de la enseñanza de las ciencias en las Facultades de Magisterio de España. Manuscrito en preparación.

Verdugo-Perona, J.J., Solaz-Portolés, J.J. & Sanjosé, V. (2016). Pre-Service primary shool teachrers' science content knowledge: an instrument for its assessment. *International Journal of Innovation in Science and Mathematics Education, 24*(2), *37-51.*

Vilches, A. & Gil-Pérez, D. (2007). La necesaria renovación de la formación del profesorado para una educación científica de calidad. *Tecné, Episteme y Didaxis*, 22, 67-85.

Contact address: José Javier Verdugo Perona. Florida Universitaria, Unidad de Educación. Carrer del Rei en Jaume I, 2, 46470 Catarroja, València. E-mail: javiverpe@gmail.com

Received: 03-08-2017 Accepted: 07-09-2018

Annex

Instrument to assess basic scientific conceptual knowledge

The contents considered are: Revolution; Galaxy; Photosynthesis; Pulmonary alveolus; Urethra; Refraction; Lunar cycle; Volcano's conduit; Sepals; Kidneys; Central nervous system; Renewable energy; Climate; Invertebrate animals; Phloem; Red cells; Liver; Chemical changes; Minerals; Organs; Virus; Zygote; Carbohydrates; Pitch; Anemometer; Flowers' vital function; Plant and animal cells; Small intestine; Filtration; Mixed colours

Some of the items are:

QUESTIONS	ANSWER OPTIONS
What is the movement of the Earth around the Sun called?	a) Rotation b) Precession c) Revolution* d) Circumference
What part of a volcano is the conduit?	a) Underground pool where magma resides temporary b) Orifice connecting volcano to the exterior c) Pipe by which magma flows up* d) Rocks and solid material formed when lava gets cold
What is the name given to the leaves that form the calyx of a flower?	a) Sepals* b) Stamens c) Petals d) Corolla
Which blood component has the function of carrying oxygen?	a) Plasma b) Platelets c) White cells d) Red cells*
What is the bending of light rays called when they pass from a fast medium to a slower one?	a) Reflexion b) Diffraction c) Attenuation d) Refraction*
What kinds of changes modify the composition of matter?	a) Physical changes b) Chemical changes* c) Biological changes d) None. Matter does not change

Received: 03-08-2017 Accepted: 07-09-2018

Instrument used to assess science process skills knowledge

The entire instrument is available at the following Web address: http://repository.up.ac.za/bitstream/handle/2263/24239/dissertation.pdf?sequence=1

Some items are:

QUESTION	ANSWER OPTIONS
In a radio advertisement, it is claimed that Surf produces more foam than other types of powdered soap. Chudwa wanted to confirm this claim. He put the same amount of water in four basins, and added 1 cup of a different type of powdered soap (including surf) to each basin. He vigorously stirred the water in each basin, and observed the one that produced more foam. Which of the factors below is NOT likely to affect the production of foam by powdered soap?	a) The amount of time used to stir the water. b) The amount of stirring done. c) The type of basin used* d) The type of powered soap used
A school gardener cuts grass from 7 different football fields. Each week, he cuts a different field. The grass is usually taller in some fields than in others. He makes some guesses about why the height of the grass is different. Which of the following is a suitable testable explanation for the difference in the height of grass.	a) The fields that receive more water have longer grass* b) Fields that have shorter grass are more suitable for playing football c) The more stones there are in the field, the more difficult it is to cut the grass. d) The fields that absorb more carbon dioxide have longer grass.
A learner wanted to know whether an increase in the amount of vitamins given to children results in increased growth. How can the learner measure how fast the children will grow?	a) By counting the number of words the children can say at a given age. b) By weighing the amount of vitamins given to the children c) By measuring the movements of the children d) By weighing the children every week*

Received: 03-08-2017 Accepted: 07-09-2018

A Biology teacher wanted to show her class the relationship between light intensity and the rate of plant growth. She carried out an investigation and got the following results.

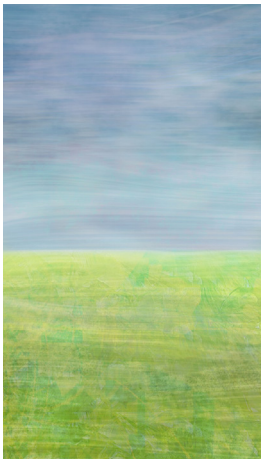
Light intensity (Candela)	Plant growth rate (cm)
250	2
800	5
1000	9
1200	11
1800	12
2000	15
2400	13
2800	10
3100	5

Table 1.5. Shows the relationship between light intensity and the growth rate of a plant.

Which of the following statements correctly describes what these results show?

a) As light intensity increases, plant growth also increases
b) As plant growth increases, light intensity decreases
c) As plant growth increases, light intensity increases then decreases
d) As light intensity increases, plant growth increases then decreases*

Received: 03-08-2017 Accepted: 07-09-2018



Reviews

Reviews

Martínez, M., Esteban, F. Jover, G. and Payá, M. (2016). *Education, in theory*. Madrid: Síntesis. 219 pp. ISBN: 978-84-9077-438-0

Place theory into controversial situations derived from practice and think, think about education "time and again, relentlessly" (p. 11): that is the aim of this book, orchestrated by four key authors in the discipline of the Theory of Education. Their work draws the reader into a reflection on education that theory may not always be as theory would suggest and yet is more in need of it than ever.

The book first looks at four major thematic pillars: learners, teachers, what is taught and learned, and finally, the educational practice. The first main section, *What Good are Teachers?* starts off with a tip of the hat to teachers and their personal influence: *Without you none of this would have happened*. The reader is invited to view teaching as a way of life characterized by personal excellence at: a) transmitting the best, b) creating unique learning spaces, and c) adding a personal touch to their students. The authors approach the reality of teaching under the premise of the latent predominance of instrumental rationality at educational institutions, which in consequence leads to the loss of the personal dimension and identity of the learners. Reflection on the teacher-student relationship reveals a bind founded on help, friendship, respect, trust, optimism, equality, listening, and reciprocity, all geared toward an eminent goal to reach in the search of an educational influence.

The second section turns to the question *What needs to be learned today?* As a result of the principle of instrumental rationality configured under the coordinates of effectiveness and efficiency, everything deemed useful nowadays is aligned as true. The authors warn about the risk of educational knowledge eclipsing everything else regarding thought and its exercise in reflection as something that basically boils down to a disdain for classical culture. However, is everything that is true also deemed to be useful? Our experience with classical culture is precisely

what lets us establish a universal continuity between distant generations. If it is the duty of education to propagate the possibilities for a fuller life, why restrict our action to the territory of the useful? What matters is not the usefulness but the worthiness; all school knowledge should be worth something to the learner.

The third section sets its gaze on the learner and the two prevailing traditions on how students are thought of: the liberal and the communitarian. Because of the predominance of the liberal perspective, they warn about the negative effects such an imbalance may have on education, particularly on its moral dimension. Although both perspectives have been out of kilter for years but have coexisted in teaching practices nevertheless, the authors make it a point to show their complementary and compatible nature. Their analysis does not lack in confrontation between these two ways of understanding education: as a civilizing process and as natural development, each with very different implications for education.

After outlining their reflections on teachers, leaners, and what should be taught and learned, the authors use the final section to discuss educational practices. Their critique revolves around three main axes that govern the world of education: a) obfuscation to fit into the surroundings, b) affirmation that educational practice belongs to everyone, and c) the excessive importance of the learner in his own education. Finally, the last chapter is followed by a brief historical pilgrimage that captures the essence on the configuration of the Theory of Education as a disciplinary, curricular field in Spain. As a whole, the sum of the chapters lends itself to an interactive, open reading, flexible for all kinds of audiences. It is especially suited for "curious, hungry minds" (p.10) who, thirsty from their curiosity and dissatisfaction, enliven and enrich the discussions that hold up the paths of education.

Laura Camas Garrido

Prendes Espinosa, M.P. & Román García, M. (Coords.) (2017). *Entornos personales de aprendizaje: Una visión actual de cómo aprender con tecnologías*. Barcelona: Octaedro. 181 pp. ISBN: 978-84-9921-901-1

The concept of personal learning environment (PLE) goes beyond a simple collection of digital tools or a platform to gather the resources used in learning processes, we must understand it considering cognitive processes, strategies and personal attitudes. The PLE is a personal representation of everything that affects our learning inside or outside the classroom, face to face or virtually.

This work reflects the collaboration of 21 authors, and it is structured in two parts. The first one places us in CAPPLE, a project funded by the Spanish Ministry of Economy and Competitiveness (EDU2012-33256) where researchers analyse the competences for lifelong learning based on the use of the PLE. Participants are students of the last year of the Spanish Bachelor's Degrees. The second part offers an expanded view through the contributions of external authors, with other visions or concretions about the PLE as well as establishing relationships with other projects or environments.

The first block, "*El proyecto CAPPLE*", is made up of four chapters that gradually allow us to deepen our understanding of the PLE and the CAPPLE project. In the first chapter, the coordinators of the book offer a general approach to the concept and introduce us to the developed project. The second chapter, from the hand of González, Sánchez and Castañeda, complements the previous one and broadens it with a review of theories and characterization of the investigations that have been previously done on the PLE. In the third place, we find a space mainly dedicated to explaining the creation and validation of the questionnaire used to carry out the research. Finally, in the fourth chapter, Gutiérrez, López, Serrano and Solano present outstanding results of the data obtained as well as a series of substantiated proposals addressed to the university institutions for the improvement of the PLE.

The second block is entitled "*Aportaciones y visiones en torno al PLE*" and allows us to deepen the concept over four more chapters, with contributions from prominent authors in the field. Salinas and Marín present the PLE through a review of the LMS, emphasizing that evolution focuses on the fact that students no longer have to adapt to a system, but

that each student configures their own space. In the sixth chapter, Aguiar, Artíles, and Rodríguez reflect on didactic and organizational approaches from the learning environments in the training of higher education teachers. We reach the seventh, where Galván and Molas focus their contribution on highlighting the relationship between PLE and digital portfolios; they do so helped by the exemplification through the project *Carpeta Digital*.

Finally, the eighth chapter of the monograph, which still belongs to the second part of the book, has a different format. Entitled "*Visiones en torno al PLE*", international authors offer individually their own perspective. Bartolomé focuses his contribution on the quality of self-regulation of learning that unquestionably throughout the book is proved as an essential element of the PLE. Ebner offers his contribution enriched by the knowledge of his participation in the PLE projects of the Technological University of Graz. Martínez contributes the most critical vision of the book by debating reflections that should not be ignored. Cabreiro is about the PLE "on the way to learning to learn". And, finally, Attwell sets his sights on the future by highlighting the keys to the relationship between the PLE and MOOCs, open education and learning analytics, among others.

In short, it is a monograph for different profiles of readers willing to go in depth on their knowledge of personal learning environments. The students or new profiles in the field will find multiple visions and conceptualizations to facilitate the understanding of the PLE, while researchers, professors or profiles that have a good knowledge of the subject will be able to deepen in a complete conceptualization at the same time that discover data and conclusions extracted from the CAPPLE national project. The variety of authors and the composition of the monograph make it an example of general interest.

Jordi Mogas Recalde