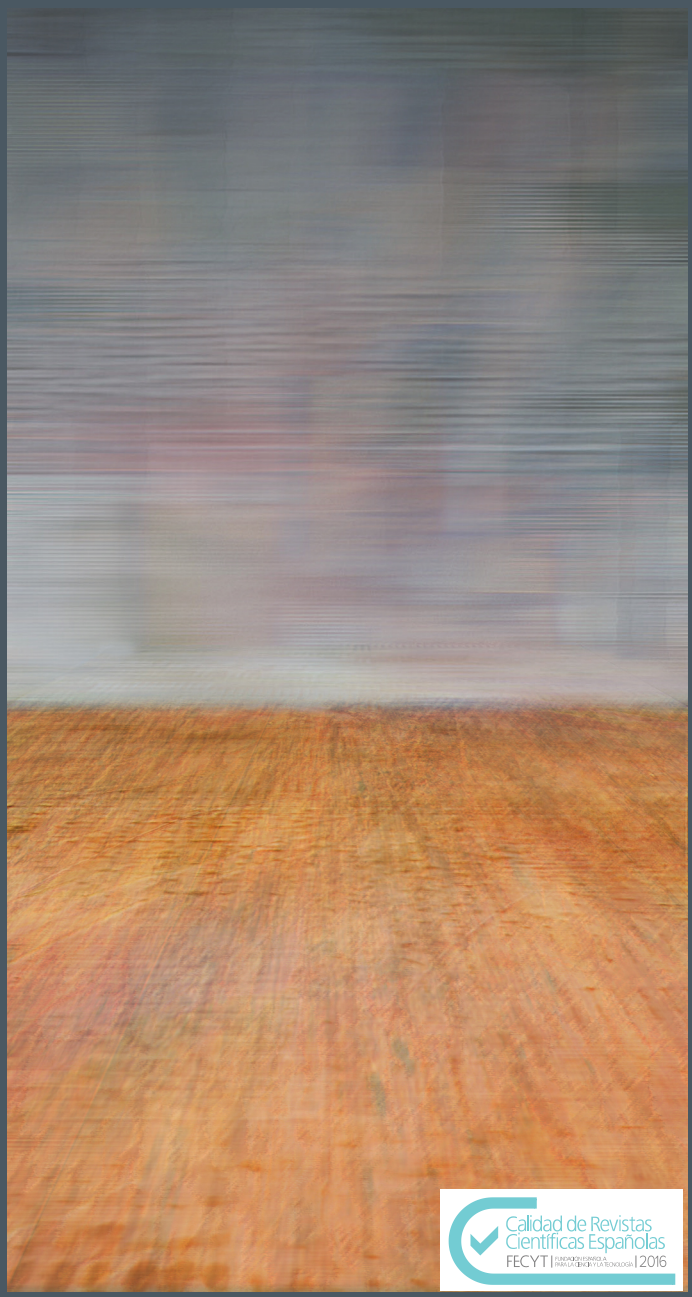
revista de EDUCACIÓN

Nº 385 JULY-SEPTEMBER 2019



GOBIERNO DE ESPAÑA
MINISTERIO DE EDUCACIÓN Y FORMACIÓN PROFESIONAL

Calidad de Revistas Científicas Españolas
FECYT | 2016

revista de EDUCACIÓN

Nº 385 JULY-SEPTEMBER 2019

revista de EDUCACIÓN

Nº 385 July-September 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: sede.educacion.gob.es
Catálogo general de publicaciones oficiales: publicacionesoficiales.boe.es

Edición: 2019
NIPO línea: 847-19-002-9
NIPO ibd: 847-19-001-3
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.

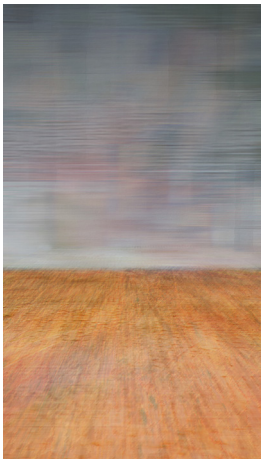
- *National catalogues*: Consejo Superior de Investigaciones Científicas (CSIC-ISOC), Red de Bibliotecas Universitarias (REBIUN), Centro Nacional de Innovación e Investigación Educativa (Ministerio de Educación, Cultura y Deporte), Catálogo Colectivo de Publicaciones Periódicas en Bibliotecas Españolas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

Implementing Governance reforms: Policy games in European university systems[1]

Implementando las reformas de la gobernanza: Las reglas de juego en los sistemas universitarios europeos

DOI: 10.4438/1988-592X-RE-2019-385-415

Karsten Krüger
Fundación Bosch y Gimpera, Universidad de Barcelona
Martí Parellada
Universidad de Barcelona. Instituto de Economía de Barcelona
Daniel Samoilovich
Asociación Columbus
Andrée Sursock
European University Association

Abstract

Since the 1990s, many European countries have reformed their higher education systems.The reforms were characterised by the disappearance of direct state control of higher education institutions (HEI) in favour of greater autonomy of universities. At the same time new mechanisms for regulation and accountability were introduced, increasing the control of institutional results.

This article analyzes the processes of structural reforms in six European countries: Austria, Denmark, Finland, France, the Netherlands and Portugal. Despite the differences between each of these countries in relation to their respective reform processes, common trends can be observed in terms of governance, human resources policy, financing, quality assessment and mergers.

National reforms and their results confirm that the governance of higher education system is increasingly complex. The state is no longer the exclusive interlocutor with HEIs, but shares it with other social actors such as: autonomous agencies

(1) This article is part of a larger study that has been founded by Cámara de Comercio de España, Conferencia de Consejos Sociales de las universidades españolas and CYD Foundation.

Received: 14-05-2018 Accepted: 15-02-2019

(e.g. funding agencies, quality assurance, research councils, and higher education councils), business and industry leaders, local and regional political actors.

Reforming higher education systems is a difficult political bargaining process, which in some cases, has led to successful reforms and in other cases to changes that were partially successful. The examples presented show that reform processes take place in waves. When a balance has been struck between the advocacy coalitions, the subsystems are likely to be highly stable over time. Only when the external and internal pressure for change is perceived as unsustainable will the advocacy coalition become unbalanced and seek a new equilibrium by means of introducing reforms.

Key words: European university system, university governance, university reforms, human resources, funding, quality assurance, mergers

Resumen

Desde los años 90, muchos países europeos han reformado sus sistemas de educación superior. Las reformas se caracterizaron por la desaparición del control estatal directo de las instituciones de educación superior (IES) en favor de una mayor autonomía de las universidades. Al mismo tiempo se introducían nuevos mecanismos de regulación y de rendición de cuentas aumentando el control de los resultados institucionales.

Este artículo analiza los procesos de reformas estructurales en seis países europeos: Austria, Dinamarca, Finlandia, Francia, Países Bajos y Portugal. A pesar de las diferencias entre cada uno de dichos países en relación a sus respectivos procesos de reformas, se observan tendencias comunes en materia de gobernanza, política de recursos humanos, financiación, evaluación de la calidad y fusiones.

Las reformas nacionales y sus resultados confirman, que la gobernanza de los sistemas de educación superior es cada vez más compleja. El estado ya no tiene la exclusividad en el diálogo con las IES, sino que lo comparte con otros actores sociales como: las agencias autónomas (p.ej. agencias de la financiación, agencias de calidad, los Consejos de Investigación, y Consejos de Educación Superior), representantes empresariales, actores políticos regionales y locales.

Reformar sistemas de educación superior es un proceso de negociación política difícil, que, en algunos casos, ha acabado en cambios substanciales, y, en otros casos, ha introducido solo cambios parciales. Los casos presentados muestran que los procesos de reformas suelen producirse en etapas. Cuando se ha llegado a un equilibrio entre las coaliciones de interés, los subsistemas son, con mucha probabilidad, muy estables en el tiempo. Solo cuando la presión externa e interna hacia un cambio se vuelve insostenible las relaciones entre las coaliciones de interés se vuelve inestable y se busca un nuevo equilibrio a través de reformas.

Palabras clave: sistema universitario europeo, gobernanza de universidades, reforma de universidades, recursos humanos, financiación, garantía de calidad, fusiones

Introduction

Since the 1990s, many European countries have reformed the governance of their higher education systems. In the majority of instances, these reforms have overlapped with other far-reaching changes in the higher education system of the continent, including most notably the Bologna Process. Structural reforms are currently characterised by a general shift away from direct state control of higher education institutions (HEIs) and a move towards increasing levels of institutional autonomy. At the same time, arms-length steering mechanisms have been introduced at the systems level and greater accountability has been required to control for institutional performance[2].

This paper deals with the implementation of these reforms in six European countries: Austria, Denmark, Finland, France, Netherlands and Portugal. It is restricted to continental West European countries that have undertaken governance reforms in higher education systems which historically have been more state-driven than their counterparts in the UK and the United States (Shattock 2014a, p. 2). The implementation of higher education policies differs from that of other public policies precisely because of the interaction between a multitude of autonomous actors (e.g. education authorities, autonomous universities, professional associations, etc.) and that it is complicated further by the diffusion of authority throughout the structure (Cerych and Sabatier 1986). Here must be added the increasing interrelation between European, national and regional policies in higher education and research (Kwiek, 2012 and Shattock, 2014b, p. 185).

This paper provides a comparative description of the reforms in the six selected countries focusing on the reforms games that various actors played. The advocacy coalition framework (ACF) developed by Sabatier and others (see, for example, Sabatier 1998; Rozbicka, 2013), suggests that policy reforms depend, on the one hand, on factors that are external to the respective policy systems (such as socio-economic developments, changes in public opinion, changes in the governing coalition, etc.) but also on tensions that characterize the political subsystem. Yet, specific

(1) The 2008 Eurydice Report, 'Higher Education Governance in Europe' advocated a clear shift from direct state regulation and control towards steering or guidance mechanisms, including the creation of new intermediate actors (such as quality assurance agencies). This coincided with a shift in the internal governance models towards "new models of managerial self-governance".

Received: 14-05-2018 Accepted: 15-02-2019

reforms depend on the advocacy coalitions that prevail in the policy subsystems, such as higher education. These coalitions are formed by actors from different organizations and institutions and from different institutional and geopolitical levels of public administration. Indeed, the ACF "explicitly argues that most coalitions will include not only interest group leaders, but also agency officials, legislators from multiple levels of government, applied researchers and perhaps even a few journalists." (Sabatier 1998: 103). The framework suggests that coalitions are organized around different types of belief, including 'deep core' beliefs (that is, fundamental values and norms) and policy core beliefs, which "are the fundamental 'glue' of coalitions" (Sabatier 1998, p.103). The ACF holds that policy subsystems (as represented here by higher education) are structured around advocacy coalitions and that when equilibrium has been struck between these coalitions, the subsystems are likely to be highly stable over time. Indeed, policy stability is only interrupted when profound changes disrupt the equilibrium (Leifeld 2013) as for instance the increasing number of enrolled students in the 1970s, which transformed elite higher education to mass higher education and, subsequently, to universal access (see Trow, 2000). Each step in this transformation required far-reaching changes in the governance of the system and of their institutions (see also Bleiklie and Kogan, 2007). But, there were a number of other key factors that contributed to this pressure for change such as changes in the graduate labour market and in the configuration of the national and international innovation systems and economic structures (Gibbons et al., 1994 and Leydesdorff and Etzkowitz, 2001) as they ventured towards the knowledge-based economy.

Studies of policy entrepreneurs (see Mintrom and Norman, 2009) suggest that perturbations in the systems often result in minor reforms that establish a new equilibrium between the coalition members, but sometimes they caused radical reforms due to the critical role played by 'policy entrepreneurs'. A policy entrepreneur is defined as a person "who takes the initiative to promote particular policy problems, identify solutions, and assemble a coalition of advocates for these solutions" (see Dougherty et al., 2013, p. 7) while policy entrepreneurship addresses the question of how a coalition develops its specific policies.

This conceptual framework, combining the coalition approach with the focus on particular initiatives of political entrepreneurs, is placed among medium-range theories, equally distant from generalizations as

well as from the specifics of individual accounts. Without ignoring the complexities of each institution, the path to such a conceptual framework emerges as an inductive conceptualization of the common elements in the national experiences, enriched by the variations that amplify the range of options for policy development (see Clark, 2004).

After a description of the main reform trends, this article provides in the third section a comparative vision of the policy games played in the six countries.

Governance Reforms in six European countries

The specialist literature detects a process of convergence in governance models towards the adoption of more managerialist models and the consolidation of universities as 'complete organisations' that is, organisations with a "well-defined identity, hierarchical structure and capacity for rational action" (Seeber et al., 2014: 7). The first central question, to which governance reforms must give an answer, is always: Who takes the final responsibility for decision and about what? The second question is, in what way the responsible persons account for their decisions? The following comparative description of the reform processes will show the difference among countries, but also the common features. It is divided in two main sections: a) the changes in the governance of HEIs; and b) changes in the HEI's environment focused on funding, human resource policies, quality assurance and mergers. This separation allows addressing adequately the rebalance between institutional autonomy and political steering mechanisms.

Emergence of new governance models

Reinforcing links with society and improving internal efficiency has been a common goal. This has been translated into changes in the governance structure of HEI by strengthening the involvement of external stakeholders, especially with industrial or business experience in governing or supervisory boards. With regard to the mode of their recruitment, there are three basic patterns: in **Denmark, Finland** and **Portugal**, universities can freely decide on external members; in **Austria**

Received: 14-05-2018 Accepted: 15-02-2019

and **France** external members are partly appointed by the university and partly by an external authority; in the **Netherlands** the appointment is completely controlled by an external authority (without excluding informal consultation).

The increasing incorporation of external stakeholders is accompanied by the reinforcement of the line management (rectors and presidents) through their appointment by the governing boards and the limitation of the influence of the university staff (academics, non-academic staff and also students) in the decision processes, even though those categories maintain a considerable degree of influence in the majority of countries.

However, the inclusion of internal stakeholders and re-balancing collegiate influence in the decision-making processes of HEIs has been subject to considerable variation among countries: in **Portugal** the most powerful body is the General Council, which has replaced the purely collegiate bodies of the University Assembly and Senate. The General Council is a hybrid body in that it has the characteristics of both a collegiate and a stakeholder body. One of its essential functions is the election of the rector or president after a public hearing of the candidate selected through a national and international open call for candidates.

In **Finland**, the current governance structure of public universities is hybrid. Professors, teaching staff and students retain influence within the collegiate bodies insofar as they make decisions as to who should sit on the governing board and how long they should stay in office. The board is composed by representatives of the professors, teaching staff, students and external stakeholders (a minimum of 40%), the latter elected by the university members of the boards. The board is then responsible for determining the overall objectives of university operations and university finances, for establishing its management principles and the appointment of the rector.

In **France**, the university council maintains a strong collegiate orientation as the number of external members in the administration council is limited to eight for a board of between 24 to 36 members. The university president is currently elected by an absolute majority of the administration council, in which, since the 2007 reforms, the external members also have the right to vote. At the second governing level, the 2013 reform reshaped collegiate decision-making processes, maintaining the scientific council and the council for studies and university, while also establishing a joint academic council formed by members of these

two councils. The academic council is responsible for staff, scientific and educational issues.

In **Austria** the influence of the collegiate body (senate) regarding institutional management tasks is very limited, mainly to set up a search commission to hire new academic staff and to validate curricula, but also to elaborate a short list for rector candidates elected by the university council, which is formed by external members. But the Senate elect between 40% and 50% of the members of the council, the other members are named by the government. It is the main governance body with supervisory functions and the competence to elect and dismiss the rector.

Denmark and the Netherlands concede more competences to the governing bodies at the expense of collegiate decision-making processes. In **Denmark** external stakeholders must be the majority in the governance boards. They are selected through a separate process by the board itself. Internal board members are elected by faculty, staff and students respectively. The board selects and appoints the rector after an open call for candidates. A special 'search committee', formed by members of the Governing board, academic councils and top management, is constituted which evaluates the candidates and presents its recommendation to the board.

In the **Netherlands**, with the University Modernization Act of 1997, the university leadership remains with the executive board. The non-executive supervisory board, formed by external members appointed by the government, took over some oversight functions from the national government, including the appointment of members of the executive board. In all six countries, however, the size of the university governing bodies has been reduced; greater emphasis has been placed on individual leadership, thereby limiting collective representation; the academic senate has also been reduced in size; and the power of collegial bodies has often been limited to academic affairs. Links with external actors have been reinforced through changes in the constitution of the universities' governing bodies. Here, a significant trend has been the move away from the election to the designation of executive roles. In the six countries, university rectors today are appointed by their respective university boards and, in some of them, the deans are also appointed by the rector.

The six countries are part of a more general trend, studied by many researchers, characterised by the adoption of a tripolar system comprising the governing board, the institution's line management and the various

academic units. The balance between these three pillars, however, varies from country to country. A primary function of the respective authority at each level is the building of trust among all parties and the establishment of a culture of collective responsibility for the success of the institution. Higher education laws in most of the countries analysed seek to strike a balance between governance regulations and the organisational autonomy of the institutions, whereby universities are free to determine their own rules of internal organisation in keeping with their academic traditions and the internal distribution of power between top and middle management. This issue is related to how intermediate executive positions in the academic units are covered, where we observe important variations. Whether middle management at universities is appointed or elected is a subject that has not been studied sufficiently. Middle management play a pivotal role: strategy is also defined at the level of academic units, their understanding of external opportunities and internal core competences as well as their commitment to the development of the whole university are critical for success. We find significant differences not just across countries, but also among institutions in the same country.

All in all, the aim is to find the right balance between the procedures of line management and the autonomy of knowledge workers as represented by university academics.

Towards new policies and steering mechanisms

The review of the specialised literature and the evidence from the six national case studies all point to a clear trend towards a greater degree of autonomy, which is accompanied by the introduction of mechanisms of supervision and accountability. Funding and quality assurance are the main mechanisms employed here, combined with the universities' more flexible human resource policies and a general reconfiguration of higher education and research systems as a result of institutional mergers.

More flexible human resources policies

Universities are human resource-intensive 'businesses'. The quality of their teaching, research and contribution to society depends primarily on

the quality of their staff and the students they attract. A substantial part of their budget is spent on personnel. Thus, autonomy in personnel matters has been one of the key aspects of the reforms under study. This includes freedom or restrictions related to hiring, remuneration, dismissal and promotion of their staff. One of the most important steps is the change of the status of the university staff from civil servants to contractual employees. This happens at different stages of the reform process.

For instance, Denmark and the Netherlands have been pioneers in the status change. In **Denmark**, the end of civil servants started in the 1960s accompanied by measures to make it more attractive to be employed on a private basis. Nowadays, limitations exist with regard to recruitment procedures and staff salaries: while the overall number of certain administrative posts is limited by law, universities may freely hire academic personnel. Danish authorities also set salary bands for senior administrative staff. By contrast, Danish laws do not specify dismissal practices; they do, however, contain general guidelines concerning promotion procedures.

In the **Netherlands**, staff has a hybrid status. Academic staff share some benefits with civil servants (such as the national pension fund) while at the same time their salary levels and related conditions are set by the national association of universities and the unions on the basis of collective labour agreements (not set by the government). Universities are free to recruit and promote personnel, but dismissal is restricted. Within the ranks of associate professor and professor, almost all appointments are on a permanent basis, while PhD students and postdocs have temporary positions. However, due to the lack of growth of the higher education sector, it is considered a less dynamic environment with fewer opportunities for young scientists to have a career in science.

In Austria, Finland and Portugal, the status change formed part of the whole reform process. In **Austria**, the reform of the labour status of the academic staff paved the way for the governance reform in 2002. Up to that reform, all university staff were employees of the federal state and the number of employees at each university, their categories and wage level were determined by the state and the respective regulation. With the reform of 2002, the universities became the legal employers of the staff and their working conditions are part of collective agreements between the association of universities and the trade unions. The reform

was not implemented to reduce staff costs, but to achieve more flexibility in human resource management[3].

Also in **Finland**, the reform of 2009 was accompanied by a change of the labour status of the university staff, who are no longer employed by the state but by the universities. Civil servant employment was replaced by a contractual relationship. Changes involved crafting a forward-looking strategy with focal areas for research and developing a new tenure-track career system for the academic faculty, thus utilising the newly acquired autonomy. Tenure track is becoming another strategic tool for allocating resources in the universities across their schools and disciplines, as it allows the integration of promising young researchers early enough by offering them proper career prospects.

In **Portugal**, the labour situation of the university staff also changes as lifetime contracts in the public administration had been abolished. The labour situation of people working in the public administration including university staff is actually more similar to the situation in the private sector. The academic staffs are contracted on a tenure track, with a five-year probation period. Formally, those universities that have converted to a foundation status have a higher flexibility in human resource management, but in fact the legally imposed convergence between public and private tracks diminishes the potential of a differentiated human resource management, even at universities with foundation status.

In **France**, university staff is mostly civil servants. The tendency of national administrations to keep close control of all public services is a major obstacle to achieving greater university autonomy. Restrictions apply to all aspects of staffing policy. Salary bands are prescribed for academic and administrative staff, and dismissals are strictly regulated for nearly all university personnel. The state imposes promotion quotas for public servants. This explains why, in the EUA Autonomy Scorecard (Bennetot and Estermann, 2017), French universities are at the bottom of the chart for staffing and academic autonomy.

---

(2) It is important to mention that the change of labour status was framed by an agreement with the trade unions to regulate the transition to the new labour status and was accompanied by state grants to manage the transition.

More funding autonomy and diversification

Funding is the area where the delicate balance that is struck between autonomy and steering is best observed. It seems common sense that financial autonomy allows the university to act more flexibly and efficiently in facing different challenges[4], but it is also commonly argued that public funding is an effective means to steer the behaviour of the university and the higher education system in terms of quality, efficiency and equity. All six countries studied have witnessed changes in their funding scheme. These transitions have been carried out in some countries independently of the introduction of new models of governance.

In the six countries concerned, the public funding of universities is implemented by means of block grants. Nevertheless, core funding is ensured by the application of different formulas.

In **Finland**, three polynomial formulas are used to distribute core funding (64% of the entire university budget), and education, research and specific strategic goals are defined. Each university's budget is established in programme contracts of four years' duration. Strategic funding constitutes 10% of the overall budget in the funding changes introduced in 2013, and this will be significant for the development of profiling.

In the case of **Austria**, the starting budget is based on a three-year performance agreement made by each university with the Ministry of Higher Education, taking development plans for higher education as the starting point.

In **Portugal**, universities used to receive public funding as a lump sum based on the formulas originally set out as a result of agreements between bodies representing universities, polytechnic universities and the government. Over the course of time, these formulas have been modified, with indicators that measure the achievement of aims being included. According to the legal framework, performance contracts could be instituted although, effectively, no use has ever been made of them.

In **Denmark**, the greater part of public financing (around two-thirds) is allocated by means of funding formulas. In order to assign this part of the higher education budget to the individual universities, what is

(3) This is reflected in the CHEPS analysis, pointing out that the universities in five of the six countries under scrutiny (France was not covered by the study) have a high level of autonomy to manage their own budget (see CHEPS et al. 2012: 42).

known as a taximeter system is employed[5]. In addition to this funding system, dialogue-based tools have been employed to give universities complementary funding since 1999. These funding instruments are a consequence of separate negotiations between the Ministry of Science and each university; they concentrate on agreed objectives and results in order to define performance contracts.

In the **Netherlands**, the basis for public funding is a lump sum as core funding, or a block grant calculated on the basis of general funding formulas. In 2012, an additional funding mechanism through performance-based contracts was set up on an experimental basis; in the first instance this was only for the 2013-2016 period, and it is to be re-assessed prior to it being transformed into a legal instrument. As regards research funding, the Netherlands is a special case as the finance is distributed among universities in the form of a lump sum. Overall only 11% of the national public budget for university finance is allocated competitively. Apart from this, competition is an essential feature in applications to a large number of additional sources of research funding, whether international or private.

In **France**, where the funding of universities is actually based on the lump-sum principle, basic funding is based on input factor. In 2008, it was complemented by output oriented budget allocation. From then on, 10% of budget assigned to teaching activities alone was based on output indicators as graduation rates and job placement of graduates. At the same time, the French government also reinforced the orientation toward project-based funding and created, in 2005-06the French national research agency. The government also increased the budget allocated to the funding of competitive research. The trend toward funding on the basis of excellence became even stronger with the state programme '*Investissements d'avenir*' launched in 2010.

Each country applies very different criteria as regards student fees and the loans and grants made available to students. Four of the six countries studied, namely Austria, France, the Netherlands and Portugal, charge fees, whereas in Denmark and Finland national and full-time EU students are exempt from tuition fees. However, in Denmark, students registered for part-time courses are obliged to pay fees, as are non-

(4) The 'taximeter system' is structured on the basis of output indicators, to a great extent in terms of the examinations that students pass. An amalgam of different rates (teaching, field work, joint expenses and building 'taximeter') are taken into account to calculate a single education rate.

EU students. Fees were introduced in Austria in 2001, shortly before the higher education system was reformed. Nevertheless, such fees are only paid by those who do not complete their studies within the time stipulated plus a further two semesters and non-EU students. Fees have existed in Portugal since as long ago as 1940, but their cost remained unmodified until it was adjusted in 1992. Five years later, the method by which the amount was calculated was changed again, the same fees being applicable to all universities; in addition, fees were tied to the minimum wage and inflation. Subsequently, in 2003 the cost of fees in Portugal was further increased, and each institution was given the right to establish their level within limits set by the government. Moreover, there is a marked trend towards introducing different fees for non-EU students. In all these countries, apart from Finland and France, students from outside the EU currently pay higher fees than those charged in the case of EU students. Since 2017, also in Finland non-EU students must pay annual fees and France plan to introduce a new policy that will enable universities to charge non-EU students fees as from 2019.

From Quality assurance to Quality enhancement

The introduction and widespread application of internal quality assurance (IQA) procedures together with the strengthening of external quality assurance (EQA) is an unequivocal indicator of universities' increasing autonomy. The European University Association's Trends 2015 report (Sursock, 2015) showed that a great majority of the 451 universities that answered the questionnaire apply IQA policies and processes and makes use of them for planning purposes and improvements within their respective institutions.

As a consequence of this expansion of internal quality processes, EQA in Europe is increasingly moving towards quality audits and institutional assessments. By and large such methods are more flexible, more closely related to the local context and guided by its requirements and, hence, pay greater respect to the need for institutional autonomy and to their ability to implement their own strategies. Another relevant aspect is the increased emphasis on enhancement rather than accountability and the implication of stakeholders and students in the tasks carried out by such bodies (Kelo, 2014).

A further development is the increased internationalisation of EQA (see Sursock, 2011) as a consequence of the greater use of international peers and the cooperation of a growing number of quality assurance agencies in individual evaluation projects (for instance, the agency for the Francophone community in Belgium and the engineering accreditation body in France worked in tandem to recognize engineering in Belgium). The recommendation on the part of the Council of the European Union that countries should permit institutions to make use of the services of any quality assurance agency to be found on the European Quality Assurance Register (EQAR) might also be mentioned at this juncture. In fact, the approach to quality assurance varies from one country to another despite the link between autonomy and the introduction of or modifications to external and internal quality assurance. **Denmark**, the **Netherlands** and **Portugal** have mainly concentrated on study programmes. All three countries have a long-established tradition of programme assessment or accreditation (for instance in the Netherlands since 1986, since the mid-1990s in Portugal). They are now all including or moving towards institutional assessment as well as programme accreditation.

In **Austria** and **Finland**, institutional quality audits have been developed that the reliability of the process of the internal quality management of individual institutions. For instance, the first Austrian quality assurance agency carried out audits of quality to guarantee the correct application of internal quality management procedures. Finland has also developed quality audits (together with reviews of specific aspects and the assessment of engineering study programmes), which are intended to provide assistance to the institutions in order to establish continuous internal quality management.

In the case of **France**, a new quality assurance agency was established in 2007. At the outset, AERES (*Agence d'évaluation de la recherche et de l'enseignement superior*) had the broadest competences among all the bodies represented in the six countries studied: it was responsible for assessing all study programmes, research groups, organisations (for example the CNRS) as well as universities. The 2013 law, however, modified the scope of this quality assurance institution (the name of which was changed to HCÉRES – High Council for Evaluation of Research and Higher Education) and restricted its capacity to evaluate research.

Regional alliances and institutional fusions

The reorganisation of national higher education systems in Europe as a consequence of the fusion of different institutions and alliances at a regional level is another fundamental feature of the structural modifications that have taken place; indeed, these reforms have led to far reaching developments in recent decades. The continent has witnessed two phases of fusion during this period: the first occurred in the 1980s and 1990s, affecting Germany, the United Kingdom, Hungary, the Netherlands, Norway and Sweden, while the more recent phase is still ongoing; the latter has affected and continues to affect Belgium, Denmark, Estonia, Finland, France, Ireland and Wales (Curaj et al., 2014).

Merger processes or a tendency towards regional alliances can be identified in four of the six countries studied in this work. Since the turn of the millennium, examples of reorganisation and department fusion leading to the creation of larger units, which seek to guarantee increased long-term stability, can be found in **Finland**, **Denmark** and **France**.

In the case of the **Netherlands**, the universities/faculties of applied sciences underwent a process of mergers in the 1990s and, at present, the approach that is being encouraged is one of increasing inter-university collaboration. Furthermore, the last ten years have seen a voluntary process of mergers taking place in higher education and research in **Denmark**, encouraged by the central government; this covers HEIs and research centres. At present, there is ongoing discussion of the opportunity of establishing a University of Denmark system, which would consist of between three and five universities and three university colleges.

Finland adopted such an approach largely for identical reasons, in other words to boost HEIs' capacity for innovation and also their international status. The HE system was considered to be both inefficient and fragmented, as well as failing to serve society to the necessary degree and failing to implement the required level of interdisciplinarity.

In both Finland and Denmark, such mergers have given rise to more professionally managed institutions. A comparable movement can be detected in **France**. It began in the early 1990s as a result of promoting regional networks through the establishment of *poles européens,* in other words the association of geographically close higher education and research organisations. Subsequently, these were converted into *pôles de recherche et d'enseignement supérieur* (PRES) by means of the 2006

law. Neither the *poles européens* nor the PRES were established on an obligatory basis; hence they failed to fulfil totally the expectation that they would restrict the negative impact of fragmentation, which is very much a feature of higher education and research in France. In 2011, a new excellence initiative (IDEX) was initiated in order to promote the idea of rewarding the most successful consortia. Moreover, to provide some balance for the uneven implementation of the PRES and growing regional imbalances that could be detected as a consequence of the IDEX, in 2013 it became obligatory for 25 regional alliances[6] known as Comue (*Communauté d'universités et établissements)* to be set up in different regions throughout the country. This law stipulated three possible organisational approaches: the fusion of already existing institutions; an association; or an umbrella organisation.

In all three countries, namely Denmark, Finland and France, the central authorities have made extra funding available in order to promote the merging of universities, polytechnics and other research institutions.

In contrast, **Austria** adopted the opposite path by removing medical faculties from universities, which resulted in their losing status in the international rankings.

Portugal has to confront two difficulties: on the one hand, a declining population; and on the other, a movement of the inhabitants from inland areas towards the littoral. This threatens the continued existence of centres located inland; they have fewer and fewer students, yet there have been no mergers as a consequence of this shift, with the sole exception of the consortia agreement signed by the three universities in existence in the Northern Region. A further case is the Universidade de Lisboa established in July 2013, after the merger of the former Universidade Técnica de Lisboa and Universidade de Lisboa.

Renegotiating governance

Despite the differences in the paths followed by the six countries as regards their aims and the application of these changes, these four areas

(5) Three types of format are legally feasible for such consortia, one of which is the Comue. However, since the majority of consortia have adopted this format, this article refers to them all by the term Comue.

in the university context – funding methods, modifications to personnel policies, quality assurance, and fusions and alliances – have made a noticeable contribution to the reformulation of HE systems.

The reported reforms confirm that, in the EU, higher education system governance is becoming more complex and involves more actors. Modifications to institutional governance have taken place in parallel to structural modifications in the higher education setting, with the setting up of research councils and quality agencies.

The state no longer maintains an exclusive dialogue with the institutions. It shares the higher education arena with other stakeholders: business and industry leaders, local and regional political actors are now involved in the running of the universities, sitting on their boards or providing them with funding instruments. Additionally, governmental steering tasks are now spread across various autonomous agencies, including those concerned with funding agencies, quality assurance, research councils, higher education councils, etc., that act as a buffer between their respective ministries and higher education institutions, and a counterbalance to the power of the ministries. For this reason, we are talking about a complex political process at different levels, which can also be observed in the reform games themselves.

One central theme to emerge from the national reform processes in the six countries has been the "renegotiation of the state-university relationship" (Kwiek, 2012: 29) striking a new balance between government steering and institutional autonomy. We observe that institutional autonomy has been strengthened in all six countries, but that at the same time the respective governments have retained control over the system as a whole (Maassen and Van Vught, 1988). There is a shift from the state control to the state supervision of universities. The increasing application of block grants, funding by contracts (or objectives) and the assignment of additional funds linked to specific objectives indicate that the state has not renounced its role in steering the system, in general, and institutional behaviour, in particular. But the state has introduced a shift in focus from "substantive policies (precisely setting what should be done and how) to more procedural policies (setting principles and aims)" (Musselin and Teixeira, 2014).

The descriptions provided of the reform processes in the six countries point to other shared features. For example, in the 1980s the need to professionalise institutional management was first debated, leading to

corresponding reforms being introduced from 1990 onwards. However, only in a few countries did these initial reforms mature into a consolidated model. Even in the Netherlands, two reform stages were needed in the 1990s to consolidate its management model. Austria, Denmark, Finland, France and Portugal had to await, at least, a second wave of reforms in the 2000s, which saw them taking steps towards greater institutional autonomy and the professionalization of their institutional management.

The reform process in each of the countries analysed has progressed at its own specific pace; moreover, these reforms clearly reflect each individual country's particular socio-economic background and culture. Nevertheless, there are a number of characteristics found in all of them. It should be stressed that the advocacy coalition approach defines three overall stages within policy making: the formulation of policy; its implementation; and its reformulation. Occasionally, the processes are (or with hindsight they seem to be) more linear and direct; but in other cases, they tend to follow a more convoluted route.

In addition, excessive emphasis on legislation often masks processes that may well have involved a large number of recurrent interactive phases at different levels. The series of reforms that took place in Austria in the last decade of the 20th and the first decade of the 21st century are revealing examples of this state of affairs. In the early 1990s, a structural reform was applied; its impact was not considered to have been far-reaching enough by some of the principal stakeholders. This resulted in the preparation of the next stage in the reform process, in which the Conference of Rectors played a prominent part. It should be emphasised that the changes in the employment status of university personnel were defined in an agreement with the trade unions in order to guarantee that the changeover to the new labour status was not so conflictive.

Comparable coalition processes can be detected in the other countries analysed. In the Netherlands, for instance, reconsideration of and feedback on its sweeping measures of reform, which drastically altered the way in which institutions acted and indeed the higher education system as a whole, brought about a range of lesser reforms. Another case of step-by-step modification of a series of legal changes can be found in France, where there tend to be frequent alterations to the way in which the adjustment process is envisaged. At some moments, research is the subject that is most prominent; at other times, education or third mission activities come to the fore. On some occasions, it is deemed

essential to develop the management of central institutions, whereas at other moments, it is considered vital to increase the role of academic staff in decision-making processes.

Changes take about two decades to bear fruit. In the 20 years that elapsed between the mid-1980s and mid-2000s, the political climate and economic context was such that it gave governments, whatever their political orientation, enough confidence to implement significant changes to their models of governance: the introduction or strengthening of competence among institutions; block grants complemented with performance-based funding, shifting from civil servant staffing models to ones based on contractual staff, the setting up of smaller government bodies with external members that appoint the rector, the consolidation of evaluation/accreditation systems – all those dimensions are present in the reform equation. The main pillars of those reforms have been constant across the countries, but the sequence in which they were introduced have been largely country-specific.

Effectiveness in the implementation of reform measures seems to be closely linked to the achievement of a broad consensus among the main actors. In the Netherlands, the 1990 reforms were based on a prior consultation process. Austria illustrates both the importance played by the Rectors' Conference as a key actor in the process and the importance of social bargaining, most notably with the trade unions, in smoothing the path for reform. In Denmark, an independent group of social actors paved the way to the 2003 reform measures, decisions that were subsequently subject to evaluation by an international expert group which led to the 2011 reform.

Another precondition for effective reform processes is that they are framed in a clear, consensual national strategy for research and higher education and that the system's priorities are clearly defined. Denmark, Finland and the Netherlands all demonstrate this link between national strategies and priorities in their commitment to developing the 'knowledge society or economy'. In Austria, the Scientific Council claimed a much stronger role of the state steering the system towards a better system performance by calling for an integrated higher education policy as well as a national research strategy.

Achieving effective implementation requires the development of a clearly-defined approach including top-down strategies alongside bottom-up initiatives. This does not rule out ambiguity; conflict is bound

to arise. It is doubtful whether substantial modifications imposed from above by ministers will achieve their objective. A successful approach requires a coalition of actors that participate at different levels of the HE system, as the advocacy coalition framework puts forward. Only through the collaboration of policy-makers (for instance at a ministerial and institutional level) is a well-defined strategy likely to be possible. By way of examples, the case of Austria and Finland show the role of the Rectors' Conference in the definition and re-definition of reform policies. In the Netherlands, the role played by the ministries concerned in the promotion of the initial reform was decisive.

In order to guarantee that reforms are successful, it is usually necessary to ensure the involvement of a grouping of political and social stakeholders, together with staff and student representatives, all of whom express their views through advocacy groups. Rather than fundamental beliefs and opinions, the driving force is very likely to be a certain feeling of urgency, powered by external factors and not purely academic concerns.

Often, modifications to the socio-economic context can lead to HE's reforms. A clear example of such a situation arose in Finland, where the reforms introduced in 1990s were equally a reaction to the shock caused by the end of the Soviet Union, the country's principal trading partner. An additional and essential external factor is that of changes in public attitudes. In Denmark and Finland, HE league tables have had a substantial impact. In both countries, these tables are conducive to the general public's awareness of the need for the existence of research-focused universities that are visible on the world stage; this has resulted in the merger of many universities. Such league tables are also a revealing indicator of the internationalisation of a country's higher education system. Nowadays, universities have to compete at a national and an international level for students, academic staff and research funding. At the same time, within the EU, the Union's strategies are increasingly having an impact on each member state's education policies. The EHEA and the Bologna process, especially the greater prominence given to quality, alongside the application of multiple reforms, has emphasised and increased the responsibilities of university management teams. Other aspects, for example a change in a country's ruling coalition or changes in the way other sectors of government are administered may result in unintended consequences for its higher education sector. Even

so, experts from both Denmark and the Netherlands underline the importance of attaining a consensus for the development of a coherent strategy for higher education and research.

Conclusion

Several clearly identifiable trends in the governance of higher education institutions have been confirmed in this study: the most prominent are the growth in competition between centres and the increased autonomy that they have been granted, although, at the same time, the higher levels of external accountability must also be borne in mind. Generally speaking, such higher levels of public accountability have implied the introduction of a tripartite system, made up by the governing board, the centre's line management, and the various academic departments. This results in the enhancing of the role of management in the universities, the inclusion of external stakeholders in the bodies responsible for governing the centre, and the need to redefine the role of collegiate bodies, with the aim of attaining a finely tuned equilibrium between these three sectors, which logically is different in each country. Simultaneously, increased levels of autonomy also show that central governments have tended to acquire an enhanced steering role by making use of, for example, the allocation of funding as a means of furthering the goals of their strategic approach.

In contrast to the trends described, some national university systems in Europe have not introduced many of the aforementioned reforms. The Spanish university system is one of them. In this system an advocacy coalition with enough power to change the logic of the university system in line with what has happened in the university systems of the countries analysed in this article[7], has never been formed. So, to highlight just two aspects, first, it is important to emphasise that universal suffrage as the governance model of universities is extemporaneous to what is found in most European countries of reference; second, unlike those countries, permanent academic staff in, Spain are civil servants.

The experiences described in Kruger, Parellada, Samoilovich and Sursock (2018) also show that the reform of higher education systems

(6) On the influence of the international environment on national university ystems, see Dobbins and Knill (2017)

is a complex political bargaining process based on advocacy coalitions, which in some cases, are driven by political entrepreneurs, and have led to successful reforms or changes that were partially successful. The examples presented show that reform processes take place in waves. When a balance has been struck between these coalitions, the subsystems are likely to be highly stable over time. Only when the external and internal pressure for change is perceived as unsustainable will the advocacy coalition become unbalanced and seek a new equilibrium by means of introducing reforms.

Reference list

Bennetot, E and Estermann, T. (2017). *University autonomy in Europe III. The Scorecard 2017.* European University Association (EUA), Brussels

Bleiklie, I. and Kogan, M. (2007) Organization and Governance of Universities. *Higher Education Policy*, Vol. 20, Issue 4, 477-493.

Dobbins, M. and Knill, Ch. (2017) Higher education governance in France, Germany, and Italy: Change and variation in the impact of transnational soft governance, *Policy and Society*, Vol 36, Nº 1, 67-88,

Cerych, L and Sabatier, P. (1986) Great Expectations and Mixed Performance. The Implementation of Higher Education Reforms in Europe. Stoke-on-Trend: Trentham books.

CHEPS, IoE and Technopolis group (2012) *Progress in higher education reform across Europe. Funding Reform. Volume 1: Executive Summary and main report*. Center for Higher Education Policy Studies. Recuperado de https://ris.utwente.nl/ws/portalfiles/portal/5141288/FUN+Vol+1+Main+report.pdf

Clark, B.R. (2004) *Sustaining Change in Universities, Continuities in case studies and concepts*. New York: Open University Press.

Curaj, A., L., Georghiou, J., Harper C. and Egron-Polak, E. (eds) (2014) *Mergers and Alliances in Higher Education: International Practice and Emerging Opportunities*. Cham: Springer.

Dougherty, K. J., Natow, R. S., Bork, R. H., Jones, S. M. and Vega, B. (2013) Accounting For Higher Education Accountability: Political Origins of

State Performance Funding for Higher Education. *Teachers College Record,* Vol. 115, Nº 1, 1-50

Eurydice/European Commission/Education et Culture DG (2008) *Higher Education Governance in Europe. Policies, structures, funding and academic staff*. Luxembourg: Publications Office of the European Union.

Gibbons, M. Limoges, C., Nowotny, H., Schwartzman, S., Scott, P. and Trow, M. [Eds.] (1994) *The New Production of Knowledge. The Dynamic of Science and Research in Contemporary Societies*. London/ Thousand Oaks/ New Delhi: Sage Publications.

Kelo, M. (2014) *QA agencies in Europe - establishment, development, trends and priorities*. Ponencia presentada en la conferencia Quality Assurance in Higher Education in Southern Mediterranean Countries: A Tool to Promote International Cooperation, Brussels, Belgium, 28 March.

Kruger, K., Parellada, M., Samoilovich, D. and Sursock, A. (Editors) (2018) *Governance reforms in European University systems.* The case of Austria (Winckler, A.), Denmark (Holm-Nielsen. L. B), Finland (Lindqvist, O.V.), France (Abécassis, A.), the Netherlands (Noorda, S. J.) and Portugal (Assunção, M. & Teixeira, A.). Educational Governance Research 8, Cham: Springer

Kwiek, M. (2012) The Growing Complexity of the Academic Enterprise in Europe: A Panoramic View. In Kwiek, M. and Kurkiewicz, A. (Eds.) *The Modernisation of European Universities. Cross-National Academic Perspectives* (27-60). Frankfurt a.M. / New York: Peter Lang

Leifeld, Philip (2013) Reconceptualizing Major Policy Change in the Advocacy Coalition Framework: A Discourse Network Analysis of German Pension Politics; *The Policy Studies Journal,* Vol. 41, Nº 1, 169-198.

Leydesdorff, L. and Etzkowitz, H. (2001) The Transformation of University-industry-government Relations. *Electronic Journal of Sociology* 4 (October 2001) Recuperado de https://www.sociology.org/ejs-archives/vol005.004/th.html

Maassen, P.A. and van Vught; F.A. (1988). An intriguing Janus-head: the two faces of the new governmental strategy for higher education in the Netherlands. *European Journal of Education,* Vol. 23, Nº 1/2, 65-76.

Mintrom, M., and Norman, P. (2009). Policy entrepreneurship and policy change. *Policy Studies Journal*, Vol. 37; Issue 4, 649-667.

Musselin, C. and Teixeira, P. N. (2014) *Higher Education Dynamics. Reforming Higher Education.* Dordrecht: Springer.

Rozbicka, P. (2013) Advocacy Coalitions: Influencing the Policy Process in the EU; in Journal of European Public Policy, Vol. 20; Issue 5; pp. 838-853

Sabatier, P. A. (1998) Advocacy Coalitions: revision and relevance for Europe; in Journal of European Public Policy, Vol. 5; Issue 1; pp. 98-130.

Seeber; M., Lepon, B., Montauti, M., Enders, J., de Boer, H., Weyer, E., et ál. (2014) European Universities as complete organizations? Understanding, Identity, Hierarchy and Rationality of Public Organisations. *Public Management Review*. Published online: 28 Jul 2014. Paper version Public Management Review, Vol. 17, Issue 10, 1444-1474.

Shattock, M. (2014a) The context of 'modernising' reforms in university governance. In Shattock, M. (Ed.) *International Trends in Governance. Autonomy, self-government and the distribution of authority (*1-13). London/ New York: Routledge.

Shattock, M. (2014b) Autonomy, self-government and the distribution of authority: international trends in university governance. In Shattock, M. (Ed.) *International Trends in Governance. Autonomy, self-government and the distribution of authority* (184-198). London/ New York: Routledge.

Sursock, A. (2011) 'Accountability in Western Europe: Shifting quality assurance paradigms', in Stensaker, B. and Harvey, L. (Eds.) *Accountability in Higher Education: Global Perspectives on Trust and Power* (111-132). New York/ London: Routledge.

Sursock, A. (2015) *Trends 2015: Learning and Teaching in European Universities.* Brussels: European University Association

Trow, M. (2000) From Mass Higher Education to Universal Access: The American Advantage. Research and Occasional Paper Serie: CSHE.1.00. Berkeley, CA: University of California, Berkeley, Center for Studies in Higher Education. Recuperado de https://cshe.berkeley.edu/sites/default/files/publications/pp.trow.masshe.1.00.pdf

Contact address: Karsten Krüger. Fundación Bosch y Gimpera, Universidad de Barcelona. Passeig Salvador Espriu 4; 3º 2ª. 08800 Vilanova i la Geltru, Barcelona. E-mail: kkruger@fbg.ub.edu

Plurilingual in society, monolingual at school: teaching practices with immigrant students

Plurilingüe social, monolingüe escolar: Prácticas educativas con alumnado inmigrante

DOI: 10.4438/1988-592X-RE-2019-385-416

Catalina Barragán Vicaria
Juan Fernández Sierra
Universidad de Almería

Abstract

The incorporation of immigrant students into our schools over the last two decades has been a challenge for the Spanish education system and for teachers, who have had to respond to this cultural and linguistic diversity, as well as to the social and personal aspects, so that all have the same opportunities. In the context of this study, even though various research works show the importance of the family's home language in learning another language, and even in academic success (Cummins, 2002; Oller & Vila, 2011), others point out how the school and the educational system avoid the student's linguistic background (Franzé, 2002; Hersi and Watkinson, 2012; Hill, Macedo and Bartolomé, 2014; Pickel and Hélot, 2014). The essential objective of this work has been to investigate the extent to which socio-familial practices and educational approaches related to the student's cultural and linguistic background have been present in the continuity and educational success trajectories of immigrant students. We draw from a qualitative study conducted in Almería (Andalusia, Spain) using interviews open to twenty-eight young immigrants of both sexes, who have reached post-compulsory education, and who have a different family language to that used at school. The data collection was complemented by interviews with especially significant teachers in the students' educational career as well as other professionals (mediator, counsellor, school director). The results show that: (1) the cultural and linguistic background of students empowers them both

Received: 18-12-2017 Accepted: 08-03-2019

personally and socially; however (2) in the school context, this is practically neglected or results in didactic disorientation regarding teaching approaches; and (3) despite this, their linguistic background gives immigrant children opportunities for personal, social, academic and work advancement.

Keywords: immigrants, academic success, languages, diversity, multilingualism

Resumen

La incorporación de alumnado inmigrante en nuestras escuelas en las dos últimas décadas ha supuesto un reto para el sistema educativo español y el profesorado que tiene que dar respuesta a esa diversidad cultural y lingüística, a la vez que social y personal, para que todos tengan las mismas oportunidades. En el marco de este estudio, si bien diversas investigaciones muestran la importancia de la lengua familiar en el aprendizaje de otra lengua e incluso en el éxito académico (Cummins, 2002; Oller y Vila, 2011), otras señalan cómo la escuela y el sistema educativo obvian el bagaje lingüístico del alumnado (Franzé, 2002; Hersi y Watkinson, 2012; Macedo y Bartolomé, 2014; Pickel y Hélot, 2014). El objetivo esencial de este trabajo ha sido indagar en qué medida las prácticas socio-familiares y los planteamientos educativo, en relación a su bagaje cultural y lingüístico, han estado presentes en las trayectorias de continuidad y éxito escolar del alumnado inmigrante. Partimos de un estudio cualitativo realizado en Almería (Andalucía, España) utilizando entrevistas abiertas a veintiocho jóvenes inmigrantes de ambos sexos que han alcanzado etapas post-obligatorias, con una lengua familiar distinta a la escolar. Esta recogida de datos se complementó con entrevistas a docentes especialmente significativos en su trayectoria educativa y a otros profesionales (mediador, orientador, director). Los resultados muestran que: (1) el bagaje cultural y lingüístico del alumnado les empodera personal y socialmente; (2) sin embargo, en el contexto escolar se encuentra relegado o ausente observándose bastante desorientación didáctica en los planteamientos docentes; (3) y ello, a pesar de que este bagaje le supone oportunidades socioeducativas relevantes.

Palabras clave: inmigrantes, éxito académico, lenguas, diversidad, multilingüismo

Introducción

Over the last few decades, migratory movements have transformed the social panorama of many countries, including Spain, which is made up of

Received: 18-12-2017 Accepted: 08-03-2019

great cultural, ethnic and linguistic diversity, although to a different extent in each autonomous community. This multicultural and multilingual reality is reflected in state schools (10.04% in the 2010/2011 course), where native and immigrant students from different backgrounds coexist. This is more evident in public schools, with immigrant students making up 12.09% of the student population in Andalusia, while in Almería province, a part of the same autonomous region and the context of this study, they make up 15.54%, of which 95.6% are schooled in public institutions. With regard to its evolution, there was a marked growth in the first decade of this century up until the 2012/2013 course, when it began to decrease slightly. In terms of region of origin, the students in Almería province come mainly from African (Morocco and Senegal), Latin American (Ecuador and Colombia), European (Romania and Ukraine) countries along with others such as Pakistan.

This new sociocultural environment has proved challenging for teaching practices, as well as in the educational research field, which tries to contribute contextualized knowledge and to build on know-how from other countries with more experience of this issue, in order to provide cues that facilitate educational improvement.

Undoubtedly, the United States is one of the countries with most experience of immigration and, in particular, the debate concerning the mother tongue and teaching in the school language; this can be seen in the multiple criticisms of the political and pedagogical proponents currently advocating English-only learning ("Only English") and the upwelling of distinct voices against the slogan "One country, one language" (Borden, 2014; Macedo and Bartolomé, 2014) with the negative connotations related to the first (maternal) languages of immigrant boys and girls. With regard to this movement, Macedo and Bartolomé (opus cit.) decry that teaching a language is not just a technical question, but rather an ideological and political question of cultural subordination; since its role is not considered when constructing the academic subjects, the students' linguistic background is therefore being ignored, which is a denial of their life histories. In this same context, Fillmore (1991) pointed out more than two decades ago how learning a second language entails the loss of the first. In Europe, Gogolin (1997) showed how teaching practice is conditioned by the teachers' "monolingual *habitus*" in monolingual nations and its role in the failure to learn second languages in multicultural and multilingual classrooms.

Received: 18-12-2017 Accepted: 08-03-2019

From a less political and more pedagogical perspective, Suárez Orozco, Suárez Orozco and Todorova (2008), in a multidisciplinary study, showed that both contextual aspects and individual differences play an important role in learning a second language, the maintenance of the first language being one of them. The importance of the family language has also been highlighted in other studies and contexts. Cummins (2002) underlines its facilitating role in learning other languages; for Thorstensson Dávila (2017) pedagogy and research on language learning should adopt the multicultural and multilingual viewpoints of immigrant students, as these would support their cognitive and linguistic growth.

In Spain, Oller and Vila (2011) point out that, when there is continuity between the school and the familial languages, linguistic skill transfer for academic purposes is benefitted. Although they recognize that it is very difficult to incorporate bilingual teaching proposals in the current multilingual context, they understand that different ways of integrating students' languages into the school environment can be found (Vila, 2006).

The advantages of the mother tongue are also highlighted by García Guerra (2007) in a variety of issues, such as those of a professional nature, by helping to negotiate with another language; or affectively, by generating self-esteem and favouring the identity of the family unit. In accordance with these approaches, Barreras, Comellas, Fidalgo, Junyent and Unamuno (2009) emphasized linguistic diversity as a richness and as an educational challenge, admitting the difficulty of managing such diversity. Martín Rojo (2010) compiled evidence on the construction of inequalities in schools and multicultural classrooms. Therefore, this management would focus on strengthening resources, the objective of which would be to educate citizens in an integrated way, so that they are respectful of both the past and the future together, on a day-to-day basis - thus assuming the universal value of difference and the certainty that linguistic richness has a personal and social worth.

However, the management of linguistic multiplicity requires both political commitment and concrete actions for it to develop; hence, it seems logical to consider the family language (or mother tongue) and other languages of students in this policy context. As García Guerra states (opus cit., p. 153): "The question would be how to articulate the relationship between the languages spoken by the immigrant students and the language of the receiving country. And this in two areas -

the relationships conditioned by the language policy and the school environment."

The lack of students' linguistic diversity in the school environment has been demonstrated in different works. Martín Rojo and Mijares (2007), in a study on the management of multilingualism in Spanish classrooms, underlined how the school tends to ignore the reality of its students, who might be bilingual or even trilingual. They also showed how linguistic diversity, if present, manifests itself in the idea that the use and maintenance of the immigrant student's language is undervalued in the linguistic market; an incompatible idea in terms of successful school integration. Accordingly, they point to a lack of sensitivity by the teacher towards the student's linguistic background and the teacher's lack of training. On other occasions we find simply foreign language teaching models, as is the case with Arabic in the proposals for Teaching the Language and Culture of Origin, because in fact, as indicated by Timaltine (1999), it is usually not their mother tongue but the standard version of that language. Indeed, schools barely consider the knowledge brought by students of foreign origin, including their languages (Franzé, 2002). Furthermore, a section of the teaching staff even perceive them as an obstacle to integration and learning. In line with this same idea, Mijares (2007) remarks how the strategies set in motion by some teachers prove that ways exist on how to structure them in the interest of greater understanding regarding linguistic diversity, in ways that allow us to value the capacities of students of immigrant origin. All of this without losing sight of the importance of learning the new language, which as Martín Rojo and Rodríguez (2016) show with the term "linguistic mutism", configures the trajectories of young immigrants as multilingual speakers, and their resulting social mobility.

In European Union countries, amongst them Spain, Parras, Rodríguez Gómez and Vale Vasconcelos (2009) point to the very limited teaching of the mother tongue through bilateral agreements with some of the immigrant population's countries of origin. Similarly, we can highlight the work by Hersi and Watkinson (2012), which shows how positive relationships between students and teachers who are sensitive to language and culture, achieve greater student involvement in the learning process, language included. However, in some countries, such as France, Pickel and Hélot (2014) reveal how the school does not consider the multicultural and multilingual competencies of the students. And in

Norway, Torbjørnsen Hilt (2017) reveals how socialization alternatives arise informing them if they can include or exclude students according to their mother tongue. The European Union therefore warns of the key role language plays in the educational process, and in terms of fairness. What is essential is that we accept and promote linguistic and cultural plurality so that it advances in a society that is ever more inclusive and open to diversity (Coste, 2014).

Methodology

The main objective of this work has been to study to what extent social-family practices and educational (curricular) approaches, have been present in the continuity and scholastic success trajectories of young immigrants in relation to their cultural and linguistic backgrounds.

An interpretative qualitative approach was chosen (Flick, 2004; Guba, 1985; Pérez Gómez, 1998) with an open, flexible and emergent design subjected to changes in the same investigative process. This methodology has allowed us to take into consideration the participants' subjective representations as well as the complexity of the relationships between people and the concern for both singular and differential aspects. We assume the contextualization of the knowledge generated, without discarding the discovery of common elements, shared by different realities.

Participants

Subject selection was carried out according to the Glaser and Strauss (1967) theoretical sampling model. Twenty-eight young people (16 women and 12 men) participated, all had continuity and educational success trajectories whereby they had been schooled in their country of origin and then enrolled in Spanish schools in the final levels of primary education or in compulsory secondary education.

In percentage terms, the nationalities of the study subjects in the province of Almería were as follows: Peruvian, 0.8%; Pakistani, 0.8%; Ukrainian, 1%; Guinea-Bissaun, 1.9%; Russian, 2.8%; Romanian, 15.5% and Moroccan, 32.6%.

All the study subjects spoke one or several languages other than the school language. These included: Arabic, Dariya, Tamazight (Berber), Romanian, Hungarian, Russian, Ukrainian, Creole, Portuguese, Urdu, Punjabi and English. All were first generation, not only from the migratory but also from the "generational" standpoint - they were the first in the family to reach these education levels and/or go to university; having studied compulsory-secondary and post-compulsory education in urban education centres in Almeria (almost all of them public, except for two charter schools). Three of these schools are in disadvantaged neighbourhoods whilst the rest are in socially integrated areas.

Instruments and Procedures

Interviews were the main technique for data collection, carried out over two school courses; this technique was chosen because we intended to give voice to the main protagonists of the educational process, the immigrant boys and girls. We conducted open qualitative interviews and non-standardized interviews (Flick, 2007) with flexible questions and some narrative stimuli (a routine clock or life experience river). Different themes were explored so that the subjects would develop them freely, by means of *in situ* "mental essays" (Stake, 1998) to motivate them to share their experiences. They were asked about their perceptions and life experience, their expectations and aspirations; about their current work, academic and/or personal reality and about the role given to their linguistic backgrounds and learning the new language in relation to their educational and personal trajectories.

To this end, we ask ourselves the following questions: How has the student's cultural and linguistic background manifested itself in their life and their school journey? What role does this linguistic background, including the new language, play in their educational and personal trajectories?

The information obtained from the young immigrants was contextualized and contrasted with the data collected through semi-structured interviews with prominent teachers who were significant in the student's school careers, as well as other education professionals - mediators, counsellors and school directors. This has allowed us to relate their teaching practice reflections to the educational experiences

Received: 18-12-2017 Accepted: 08-03-2019

of the young people themselves and, therefore, achieve greater depth and explanatory potential.

For the analysis and categorization process, the guidelines of Flick (2007), and Taylor and Bogdan (1987) were followed. It was a recurrent process in the different phases of the study. Firstly, we traced the emerging themes from the information obtained from the interviews of the young immigrants, contrasted by the contributions of their teachers and other educational agents, which, in a choral fashion, provided a complementary image; making it possible to triangulate with other data from informal observations as well as negotiating the reports with teachers and education professionals.

Results

From the data analysis, the linguistic question emanates as a central axis in the educational and personal trajectories of the young immigrant participants, from which the following important points emerge: The fundamental role played by the mother tongue and other languages in the recognition and socio-familial position of the students; the school's ignorance regarding their linguistic background as well as the didactic disorientation in teaching practices; and the academic, social, personal and work possibilities that this multilingual competence opens up for the immigrant students.

The mother tongue in the socio-familial context: empowerment and usefulness

The young people in this study know and use other languages other than the one used at school, to interact with their families and other members of the community, including their peers, using it differently and diversely depending on the contexts and situations in which they are immersed at any time.

Therefore, first of all, we encounter the primary role played by the familial language when communicating amongst family members, mainly with the parents, as we can observe in the following statements: "With my

Received: 18-12-2017 Accepted: 08-03-2019

mother and father [I speak Moroccan]. "At home, Arabic is spoken" (Majda); "My father says: At home, speak Arabic, otherwise you are going to forget it. But often at home with my sister, nothing but Spanish "(Amal); "At home I speak Ukrainian always, but if [my mother] speaks to me in Spanish, then I speak in Spanish. With my father no, only Ukrainian." (Sveta)

This communication in their mother tongue and the other languages present in their community also occurs with other relatives such as cousins and siblings etc; offering us a picture of multilingual youth. For example, Naima from Morocco: "Well [I know] French, Arabic and Spanish, and English; written, spoken [Arabic]... ". This recognition leads them to change linguistic registers depending on the context and the interlocutors: "Maybe with my cousins I speak Spanish, which they understand perfectly, or Arabic because they understand both." as Malak told us. Or Natalia from Romania:

> With my mother I speak Romanian, with my brother and my sister, Spanish, and when I go out I speak in English... I perhaps speak Hungarian with my brother, because my brother is Hungarian, because my father's family speaks Hungarian, it is on the frontier, a little lower, near the border with Hungary. There are Hungarian schools there.

The communicative exchanges in the familial, social and affective context, which are more or less spontaneous, facilitate learning the language in a natural way, because, in many cases, the skills for speaking the host country's language are very basic, linked to work situations, especially for the father: "With my father, Arabic, because we have to talk to him, because he doesn't understand, the poor thing!" (Malak). Whereas mothers, especially if they do not work outside the home, usually have less knowledge of Spanish. This is not to say that they have no interest in learning it but have fewer opportunities to do so.

> My father, he got here and at work he didn't know anything. It was very difficult, but he learned it (...). My mother, a little more difficult than us, because we are younger..., but she did learn. (…). It was at home, and talking to people, but classes, no. (Nicolae).

This situation implies, from the affective point of view, the necessity of using the mother tongue, as well as interest learning it or retaining it

Received: 18-12-2017 Accepted: 08-03-2019

in different ways: "Let's say I'm there, with films (...) which are in Arabic with subtitles, you read in Arabic and so you don't forget "(Majda). Being aware of the new language's influence:

> I talk to my parents, to my friends, to everyone. To practice the basics so I haven't forgotten it. But the accent is already noticeable. Since I have Spanish, when I speak in Ukrainian and Russian, I notice it's a bit strange. (Mariia).

This occurs not only with the familial language but also with other languages used in the community, as in the case of those coming from Guinea-Bissau:

> The 'Creole' will never leave me..., I speak it with my friends, with my father, with my boyfriend. It is there, I speak Spanish and speak "Criollo", but Portuguese, as there are people from my country who don't know Portuguese..., well, you don't have no one to talk to and you forget things (África).

The colloquial language is the language of their affections, which identifies them and gives their family the feeling of belonging, as well as cultural and origin references. This results in the first language acquiring great value within the family, which helps them to maintain the affective and relational bonds that characterize this context.

Furthermore, knowing the languages of their community, of their country of origin, gives them great self-esteem from a personal standpoint, as we can see with Abdul, a young person from Pakistan:

> I can speak English, the Spanish I'm talking to you, Urdu, Punjabi, which is the language of the city; the city, no, of a community called Punjabi, and in that community, we speak Punjabi and Urdu, which is similar to Hindi, what they speak in India.

The knowledge of two or more languages, the familial one and that of the host country, raises them to a position of recognition in the family context, being able to act as translators and mediators between their family members, other people from their country of origin and those of the host country; which, as in the previous case, leads to a certain level of social and familiar empowerment: "I help in my father's internet booth, with other people from my country."

Received: 18-12-2017 Accepted: 08-03-2019

Salma is in a similar situation, becoming indispensable for the daily family chores, "paperwork... For my parents, my brothers... To register them... if they want to go to the hospital, I have to go with them, I help outside and inside. " Salma's role remains a great support for the family, and she even manages to work following her professional training.

This mediating role places them in an advantageous position because, indirectly, it leads them to feel that school learning is valued by their family and thus, so is the school. This leads them to appreciate their studies and continue on their school journey, despite the difficulties; in the same way as their mother tongue, for the value it has in helping the life of their family and their community.

> [I speak] Romanian, English, Spanish, a little French and... My parents know Hungarian, my parents are Hungarians and I am thinking about learning Hungarian, the more languages the better (...) It depends because I'm thinking of at least the basics, speaking and all that, but in Hungarian, there are quite difficult letters and I'm not considering writing it (...) but I would like to understand, to be able, for example, to make a trip alone... That's why it's good to know any language. (Nicolae).

We observe how these boys and girls are aware of the power that knowledge of the familial language gives them along with the language of the host country and other languages of the country of origin; In other words, how learning languages expands opportunities in different contexts, beyond the school, such as the social and work context, converting this linguistic capital into symbolic capital that lifts them towards a greater academic journey and an advantageous social position (Martín Rojo and Rodríguez, 2016). This makes them explicitly want to improve the knowledge of these languages, fixing on specific actions with this aim:

> I don't know if I'll have any chance of staying here [The Official School of Languages]. like my brother..., He did a test, an exam and for six years he's been put in fifth grade, and he didn't study (...) I don't know, maybe you need someone to push you because, at first, you're afraid to do something, always maybe you want to do something and you don't know if it'll go well or badly... (Malak).

Amal, like Malak, had the same idea: "I'm thinking of signing up for Arabic at the School of Languages." Her sister had already done so, and now, in addition to university studies, she has an advanced level of Arabic that is very useful in her work.

The mother tongue in the school context: between relegation and didactic disorientation

Despite the value given by the study subjects, and their families, to the mother tongue and other languages in the community, they are practically relegated to the family environment only: "We speak Arabic at home. 'Berber'. Only speaking "(Ilham). No explicit manifestation from the participants to their first language in the school environment, despite the anguish of first arriving in a place, and an unknown group that deprives them of being able to understand and communicate with others, as Majda expressed: "You don't understand anything so you cut yourself off. Each in their place. "

In other situations, we observe certain contradictions in the feelings of young immigrants:

> I am from Transylvania, I was born there, but I go more for Hungarian, because my parents are Hungarians, they speak Hungarian and I grew up speaking Hungarian. And Romania is another thing, and the children who are born like this, with Hungarian, they are taken to another institute which is half Romanian-Hungarian and they learn both languages. (...) I, when asked I say that I am Hungarian and I don't care (...) Like it or not, you have to accept it and that's that (...). And then, if they give me Spanish nationality, I'll say I'm Spanish (...). For me, I don't like people I have seen who say that you are... I do not know what..., I think you just have to accept, white, black or coloured whatever, or the language we speak, we're all equal, only that some come from a different part of the world, and that's it. (Derek).

Testimonies like this place students in different circumstances when building their identity, in relation to the contradictory attitudes encountered in the new society. Perhaps, unconsciously, society and teachers do not value the student's linguistic background in the same way,

Received: 18-12-2017 Accepted: 08-03-2019

when it comes to languages such as English rather than other languages from economically less developed countries. This perception would be conditioned by questions of power and hierarchical relationships that are more political, economic and social rather than linguistic (Moreno Cabrera, 2000), which is also reflected in the faculty's mastery of foreign languages. However, this knowledge can be very useful for students from countries where one of these languages is official and, thus, used in their schooling; as noted in the case of Abdul, "I have studied in an English school, we had mathematics in English, and everything, that's why I'm much better at English." Also, that one of his teachers thinks it has helped him in his studies at the Institute, "Yes, I think it did [the English helped]."

Conversely, when other mother tongues are taken into consideration, proposals for positive action appear to incorporate them into the school, such as that carried out in one of the institutes attended by some of the study subjects. The teacher declared "But I, because of my ignorance of your culture, of its make up, I propose that there be a bilingual teacher for that class (...), a teacher [certified] by the Official School of Languages. " (Teacher. 1).

However, it was not positively valued, and was rejected by almost all the Moroccan students:

> It was a failure (...) I thought that these children's learning, their integration would be faster. At least, it was a person who could communicate with them better, because I always communicated with gestures, to help them (...). I don't know [I thought that] if someone came who knew their language, for these children who are so smart and so eager to work, with a person who can talk to them and open the way for them... The first year did not go well (...). Because first of all, of course, maybe this teacher came in thinking that he could give a class in Arabic..., he lacked awareness of where he was. Then he had problems because the children as far as they didn't understand what they were doing, they didn't understand him. They now didn't understand in Spanish or Arabic. So immediately the class became very conflictive. (Teacher. 1).

The failure of this proposal leads us to think about the importance of incorporating these actions by analysing the viewpoint of these boys and girls, and the consequences for their vital projects, in the way they are

perceived by others and perceive themselves and in their position in society. Therefore, it is vitally important in teaching to approach and recognize the cultural backgrounds of the students, including the language (s) from the subject's lived culture. In short, avoiding stereotypical approaches to groups and their culture –to which a language is attached– as something codified and static. In contrast, to seek proposals coming from the relationship between the people they live with in these educational spaces and to investigate their possibilities so that each finds a place to grow, in a singular way, with their cultural references, relating them to those of others. This means adopting a multicultural and multilingual perspective on teaching, a task that is not easy for teachers, because, as Gogolin (1997) points out, in the countries of our environment, educational practices are conditioned by a "monolingual *habitus*".

The linguistic background of students: socio-educational opportunities

Although no statements have appeared regarding the use of the familial language in the school, it has been done timidly with other languages known to the students, some from their schooling in an official language used in their country of origin, other than the language spoken by the family. This has been an advantage in the support they have been able to receive in the teaching and learning processes of other curricular areas.

> The first year of teaching computing, as I didn't mind him answering the exams in English, they were given in Spanish and he answered in English. But this helped him [to learn Spanish] little by little. So, in the third quarter he said: Now, I will try in Spanish. And this helped him to use the language better. Well, then, if the teacher helps you because you don't mind using two languages (...), they begin to realise that you can proceed. The bad thing is when they have difficulties and you say: No! You have to do it this way! and not be flexible. (Teacher 2).

This teacher has been able to utilise some of Abdul's linguistic background; he was previously schooled up to secondary education in Pakistan, where English is an official language. His experience leads him to reflect on this issue in schools and the difficulties involved in incorporating it:

Received: 18-12-2017 Accepted: 08-03-2019

> So, I think it can be extrapolated to other kinds of situations. What happens in the case of English is clear, but if they come with another language other than English, it is going to be very difficult for the teacher to know it. (...) French too. But you know that, in Morocco, there are also many who don't know French if they have not been in school. Especially with Moroccan Arabic, which is not [standard] Arabic - then it will be very difficult for teachers to know it or allow the student to do things in that language. Then there is the language barrier... If this is not resolved quickly, if they are not able to quickly say: I have learned Spanish; Well, maybe they'll stop studying. (Teacher 2).

Like Abdul, Nicolae values the advantages of knowing other languages, in this case English:

> I came here three years ago, almost three years, It was very difficult, I arrived a week before the course, and knew nothing, no Spanish or anything. But what was lucky was that I had an English teacher, and if I had doubts or anything, I asked in English and he explained it to me and everything.

From the students' linguistic diversity, one perceives a positive attitude towards learning other languages in most of these young people. This attitude sometimes leads to school and academic trajectories linked to language studies: "English Philology studies, English teaching, tour guide, translator and these types of things. It's just that I'm very good at English "(Marcus). Hence his commitment and motivation to learn a language will lead him to read "books in English for the vocabulary." His teacher and tutor corroborate the effort and, at the same time, facilitate learning the language:

> In the case of Marcus, it's that he has a breath-taking ability for languages, both English and Spanish. In English, he made progress..., I've had few students in twenty-four years who've made such great progress as I have seen with him, with a certain facility. (...) then Marcus asks you, ' And what's that? ' (Teacher 1).

Marcus is not the only case showing this interest in languages. Kateryna is also inclined to linguistic studies, juggling various options,

although conscious of the economic limitations, she turns it down for something she can do in the city where her family reside:

> The truth is that, from the beginning, I wanted to do interpreting and translation in Granada, but I also thought that being so far away... really my parents cannot... I don't have any type of help (...). So, it would be more difficult for me to be in Granada than here in Almería. When we went to the university, we went on a visit, just that the guy who explained everything to us said he had studied English Philology.

Some girls take advantage of their knowledge of other languages, in which they obtain good grades, to combine their multilingual learning, for example, Bouchra, or her sister Amal:

> I like languages. Because I like learning languages. Arabic, Spanish and here I learn English and French. [Do you go somewhere to learn languages?] No, but I signed up for French next year. English as well. I don't know, I like them both. [She's asked about her results] English 9, French 7 and 8. Yes, in primary school, nothing but English. (...) I signed up this year for French but... I'm doing the level test to see if they put me in the second or third (...). That's what the teacher told me. She's going to help us, and she's going to test me and such. I've done French since I came here. I also wanted to take English but the teacher told me that with the baccalaureate they put you in at the intermediate level. Well, why get put in the first level, at the basic level, if I can be in the middle? [You can only be in one, right?] The first year. The second you can do more. [You are very well informed]. [Laughs] I've read it. I wanted to get into several, in French, in Arabic, but the first year you can only choose one.

In some way, we might think that her experience learning several languages has been valued positively, enhancing high self-esteem, because with the effort and work, she has been successful. As some of the boys and girls in the study recognize: "I learned very quickly because I was interested and wanted to study" (Abdul); "Many [Hours studying]. Because well, at the weekend I sit in the chair at night and get up at six or seven in the morning "(Amin). This is supported by the perception the counsellor has: "Self-conception and self-esteem that, once these levels have been

passed, help value them as a field of greater possibility." Taking into account the value given to knowledge and mastery of various languages in the social, political and economic spheres, these young people realize that they are linguistically richer than their monolingual peers and professors. So, it seems logical that some have seen their future in language learning. Like Abdul: "Well, I speak three or four languages. Also French a little, I can talk to people, for example, when I'm in the internet booth."

For the participants in this study, having one or several languages other than that used in the school, once the obstacle of the learning the language being taught, it turns into a social and work advantage; this is commented on in the explanations given by one of Amin's teachers:

> He's figuring out his own life. Now he's working because he's a real fighter. He's been working in a lawyer's office as a translator. I think he's done other jobs, I don't know about the police or the Civil Guard. He is now working in a company... [Computer Science]. (Teacher 1).

For Abdul, his good mastery of English has gained him the Professional Training of Higher Degree of Informatics: "... I'm in London and I want to talk to people, I can talk and I can understand well. I have a level that's fine. " His teacher refers to this:

> [Abdul], for example, when he was in secondary school, as he still didn't speak very well because he had learned very quickly, he had some problems with the language, so he went to the library, grabbed an English dictionary and translated things word by word to improve his language. (Teacher 2).

His knowledge of another language, English, was erected as a bridge to learn Spanish, it also led him to make future plans in other countries to conclude his studies:

> One more year and then three months of internship. [In companies?] I don't know, but I have an idea to go abroad, to London or in the Union... what's it called, the UK. [Why London?] Because I know English better, better than Spanish, because in my country there is another national language, Urdu, but as I have studied in a school... (Abdul).

One can observe how several participants directed their attention to academic and professional trajectories linked to their knowledge of languages. Bouchra, with university studies linked to languages, studied Arabic in official language schools, which has allowed her to get a job where mastery of several languages is necessary, in particular those of immigration countries in our region, such as Morocco; the same as has oriented her sister towards university studies of this type because it offers her more job opportunities: "I don't know, I'm thinking, I want to go to college. Yes, I'm thinking, but I'm not sure yet. My sister tells me to do translation and interpreting. " (Amal).

Conclusions

In conclusion, we have synthesised the results of several important ideas, considering that the qualitative paradigm in which this study is located aims to present a description and interpretation of reality, not generalizable, but that allows the reader to understand and transfer (Guba, 1985).

First of all, we can highlight the relevance of the mother tongue and/or languages of the country of origin, in terms of the family and the community of the young people. This is because, on the one hand, it facilitates communication in the family, and hence their affective relationship; since language, as with the boys and girls' other knowledge are an important part of their life experience, allowing them to develop the social and cultural competencies linked to the contexts to which they belong. At the same time, these situations create spaces to develop their identity and to establish a relationship with it, and from it to other groups. In this sense, it becomes an important link from which to establish new relationships and learning in the new place, mediating between both contexts, the family and the host country, which raises them to a privileged position at the heart of the family, which, in turn, entails recognising school learning and prompts them to continue their studies. Lahire (1995) refers to this as the "l'intégration sociale et symbolique de l'expérience scolaire" (1995, p. 278). These circumstances lead to the recognition of school learning, which motivates them to continue studying.

Secondly, we find a lack of recognition by teachers of the linguistic and cultural background of immigrant students, as has been noted in

other research (Franzé, 2002; García Guerra, 2007; Martín Rojo and Mijares, 2007; Pickel and Hélot, 2014). But when we find some sensitivity from teachers towards the students' cultures and languages, they develop educational practices for teaching the mother tongue from their standpoint as teachers in a framework of monolingual reference, which, to some extent, conditions both the teaching of the second language and other languages (Gogolin, 1997).

Thirdly, it is easier to consider the linguistic background of the students in the school when they are official languages of the students' countries of origin, taught in the Spanish educational system, with some of the teachers able to use it to facilitate academic learning. However, this does not happen with some mother tongues because, as Oller and Vila (2011) indicated, it is difficult to have teachers who understand all the students' languages. In any case, this should not be obstacle to finding other alternatives to make students feel that their linguistic background is valued, including their first language (s) and the school language.

Fourthly, the linguistic background opens up educational (and work) opportunities embodied in the students' educational trajectories with academic and professional itineraries linked to a knowledge of languages; given that all of them also achieved good mastery of the new language. This is reflected in the multiple professional profiles and studies linked to languages, converting their linguistic capital into symbolic capital (Martín Rojo and Rodríguez, 2016) raising them to a social position of greater status and the upward social mobility that opens up the possibility of continuing post-compulsory studies; as was the case for the subjects of this work.

Fifth, the intergroup and intragroup differences encountered, question the static and codified viewpoint of the reference groups, breaking with the homogenized view of the "immigrants" as a collective, as well as the cultural and linguistic groups to which they are ascribed –by the society and the teachers– in relation to their possibilities of continuing studies and having educational success.

Finally, a reflection on the educational implications. There is a need to adopt a multicultural and multilingual perspective to teaching practices, which takes into account the students' backgrounds at the same time as learning the new language well. This involves revising certain ideological assumptions, to overcome them and make implicit the monolingual framework in which we move, not only educationally but also politically

Received: 18-12-2017 Accepted: 08-03-2019

and socially. In this way, recognizing the multilingual and multicultural diversity of schools and society, as well as the consequences of the actions taken to offer students greater educational and employment opportunities.

Bibliographical References

Barreras, M., Comellas, P., Fidalgo, M., Junyent, M.C. y Unamuno, V. (2009). *Diversitat lingüística a l'aula: Construir centres educatius plurilingües*. Barcelona: Fundació Jaume Bofill; Vic: Eumo, Recuperado de http://www.fbofill.cat/intra/fbofill/documents/publicacions/501.pdf

Borden, R. S. (2014). The English Only Movement: Revisiting Cultural Hegemony. *Multicultural Perspectives,* 16 (4), 229-233. doi:10.1080/15210960.2014.956607

Coste, D. (2014). Multilingualism and the challenges of education. In P. Grommes and A. Hu (eds.), *Multilingualism Education. Policies-practices-language development, Hamburg Studies on Linguistic Diversity* 3, (15-32). Amsterdam, The Netherlands: John Benjamins Publishing Co. doi: 10.1075/hsld.3.03cos

Cummins, J. (2002). *Lenguaje, poder y pedagogía.* Madrid: Ministerio de Educación, Cultura y Deporte – Ediciones Morata.

Fillmore, L. W. (1991). When Learning Second Language Means Losing the First. *Early Childhood Research Quarterly,* 6, 323-346.

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

Franzé, A. (2002). *Lo que sabía no valía. Escuela, diversidad e inmigración.* Madrid: Consejo Económico y Social.

García Guerra, M. (2007). *La Razón mestiza. Agenda Intercultural.* Madrid: MEC-CIDE.

Glaser, B. G. & Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Chicago: Aldine

Gogolin, I. (1997). The "monolingual habits" as the common feature in teaching in the language of the majority in different countries. *Per linguam, A Journal of Language Learning,* 13 (2), 38-49. doi: https://doi.org/10.5785/13-2-187

Guba, E. (1985). Criterios de credibilidad en la investigación naturalista. En J. Gimeno Sacristán y A. Pérez Gómez, *La enseñanza: Su teoría y su práctica,* (148-165). Madrid: Akal.

Hersi, A. A. & Watkinson, J. S. (2012). Supporting Immigrant Students in a Newcomer High School: A Case Study. *Bilingual Research Journal,* 35 (1), 98-111.

Lahire, B. (1995). *Tableaux de Familles* [Hautes Études]. Paris: Gallimard - Le Seuil.

Macedo, E. & Bartolomé, L. I. (2014). Multiculturalism Permitted in English Only. *International Multilingual Research Journal,* 8 (1), 24-37. doi: 10.1080/19313152.2014.852426

Martín Rojo, L. (2010). *Constructing inequality in multilingual classrooms.* Berlin: Mouton de Gruyter.

Martín Rojo, l. y Mijares, l. (2007). "Solo en español": una reflexión sobre la norma monolingüe y la realidad multilingüe en los centros escolares. *Revista de Educación,* 343, 93-112. Recuperado de http://www.ince.mec.es/revistaeducacion/re343/re343_05.pdf

Martín Rojo, L. y Rodríguez, L. R. (2016). Muda lingüística y movilidad social. Trayectorias de jóvenes migrantes hacia la universidad. Discurso & Sociedad, 10 (1), 100-133. http://hdl.handle.net/10486/675924

Mijares, L. (2007). *Aprendiendo a ser marroquíes. Inmigración y escuela en España*. Tesis Doctoral editada por el servicio de publicaciones de la Universidad Autónoma de Madrid.

Moreno Cabrera, J. C. (2000). *La dignidad e igualdad de las lenguas. Críticas de la discriminación lingüística*. Madrid: Alianza Editorial.

Oller, J. y Vila, I. (2011). Interdependencia entre conocimiento de catalán y castellano y efectos de la lengua familiar en la adquisición de las lenguas escolares. *Cultura y Educación,: Revista de teoría, investigación y práctica,* 3 (1), 3-22. doi: 10.1174/113564011794728588

Parras, A., Rodríguez Gómez, F. y Vale Vasconcelos, P. (2009). *La integración escolar del alumnado inmigrante en Europa. Medidas para fomentar la comunicación con las familias inmigrantes y la enseñanza de las lenguas de origen del alumnado inmigrante*. Bruselas: Eurydice. Recuperado de http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/101es.pdf

Pérez Gómez, A. (1998). *La cultura escolar en la sociedad neoliberal*. Madrid: Morata.

Picket, T. & Hélot, H. (2014). "Because it is my life, and I'm the one who makes choices" - Newcomers in the French education system and career guidance. What about their multilingual competence? In P. Grommes and A. Hu (eds.), *Multilingual Education. Policies-practices-language development, Hamburg Studies on Linguistic Diversity 3,* (161-180). Amsterdam, The Netherlands: John Benjamins Publishing Co. DOI: 10.1075/hsld.3.11pic

Stake, R.E. (1987). *Investigar con estudios de caso.* Madrid: Morata.

Suárez Orozco, C., Suárez Orozco, M. M. & Todorova, I. (2008). The Challenge of Learning English. En Autores, *Learning a New Land: Immigrant Students in American Society* (146-166). Cambridge, MA, USA: Harvard University Press.

Taylor, S.J. y Bogdan, R. (1987). *Introducción a los métodos cualitativos de investigación*. Barcelona: Paidós Ibérica.

Tilmatine, M. (1999). La importancia de la lengua y cultura de origen en la integración de las comunidades beréberes en Europa. En A. Franzé y L. Mijares (Eds.), *Lengua y cultura de origen: Niños marroquíes en la escuela española* (143-156). Madrid: Ediciones de Oriente y del Mediterráneo.

Thorstensson Dávila, L. (2017). Newly arrived immigrant youth in Sweden negotiate identity, language & literacy. International Journal of Inclusive Education, 21 (6), 1-11. doi: https://doi.org/10.1016/j.system.2017.04.001

Torbjønsen Hilt, L. (2017). Education without a shared language: dynamics of inclusion and exclusion in Norwegian introductory classes for newly arrived minority language students. *International Journal of Inclusive Education,* 21 (6), 585-601. doi: 10.1080/13603116.2016.1223179_

Vila, I. (2006). Lenguaje, escuela e inmigración [Monográfico: Lengua e inmigración]. *Cultura y educación. Revista de teoría, investigación y práctica,* 18, (2), 127-142.

Contact address: Catalina Barragán Vicaria. Universidad de Almería, Facultad de Ciencias de la Educación, Departamento de Educación. Ctra. Sacramento s/n. La Cañada de San Urbano 04120 Almería. E-mail: cbarvi@ual.es

Received: 18-12-2017 Accepted: 08-03-2019

Inferences of textual comprehension and genres: A comparative study between Costa Rica and Spain[1]

Las inferencias de comprensión textual y los géneros discursivos: Un estudio comparado entre Costa Rica y España

DOI: 10.4438/1988-592X-RE-2019-385-417

Fernando Guzmán-Simón
Celia Moreno-Morilla
Universidad de Sevilla
Isabel Gallardo Álvarez
Universidad de Costa Rica
Eduardo García-Jiménez
Universidad de Sevilla

Abstract

The various theoretical models of reading comprehension have revealed the complexity of the cognitive processes involved. Readers need to construct the meaning of what they are reading, to check what they are reading against their recall of related information, and to build a coherent frame of reference or *reference model*. This reference model develops from the *text-base*, as well as from the integration of this text-base with the reader's previous knowledge, derived from those readers' experience, their worldview and their understanding the text's genre in the text-base. Our research analyses the construction of

(1) This research has been carried out within the framework of the projects *Inferences and self-regulation in digital and printed reading in Costa Rica and Spain* (Pry01-1037-2015), funded by the Ministry of Education of Costa Rica, and *Improvement of reading inferences and self-regulation in secondary school students in Costa Rica through training in teacher evaluation* (UCR-US) (EDU2012-34000), funded by the Development Cooperation Office of the University of Seville.

inferences throughout different genres within a sample composed of 3,199 Secondary Education students in Costa Rica and Spain. To this end, TECOLEIN, a test aiming to measure students' level of inferential reading comprehension, was designed and validated. TECOLEIN was composed of four types of inferences in three different genres. The data obtained were analysed using Paired-Samples T-Test and ANOVA and, at the same time, a two-stage cluster analysis was performed. The results obtained in TECOLEIN have allowed, firstly, to observe a gradual progress of the inferential reading comprehension among students aged 12, 14 and 16. Throughout secondary education, the students develop strategies that integrate into the reading comprehension both the knowledge built from the text, and that based on long-term memory. Secondly, the joint analysis of the answers offered by the students to questions of inferential comprehension related to different genres reflects the existence of different degrees of difficulty in the comprehension of a text. Finally, difficulties of reading comprehension in Costa Rica and Spain present scaled results, since both countries present a similar difficulty associated with a type of inferential comprehension.

Key words: reading comprehension, literary genre, Reading comprehension assessment, high school students

Resumen

Los distintos modelos teóricos sobre comprensión lectora han permitido comprobar la complejidad de sus procesos cognitivos. El lector ha de buscar la congruencia de sentido de la lectura, validar los procesos de recuperación intencional de la información y construir un modelo de referencia coherente. La construcción de un *modelo de referencia* partió de la construcción del *texto-base* así como de la integración en el texto-base de los conocimientos previos del lector, relacionados con su experiencia, visión del mundo y el género discursivo del texto. Nuestra investigación analizó la construcción de inferencias a través de distintos géneros discursivos con una muestra compuesta de 3.199 estudiantes de Educación Secundaria en Costa Rica y España. Con tal fin fue diseñado y validado TECOLEIN, prueba que tiene como objeto medir el nivel de comprensión lectora inferencial de los estudiantes. TECOLEIN estuvo compuesto por cuatro tipos de inferencias en tres géneros discursivos distintos. Los datos obtenidos fueron analizados, por un lado, con la prueba t para muestras relacionadas y ANOVA; por otro, un análisis de clúster bietápico. Los resultados obtenidos en TECOLEIN han permitido, en primer lugar, observar una graduación de la comprensión lectora inferencial en los estudiantes de 12, 14 y 16 años. Los estudiantes desarrollaron en la Educación Secundaria estrategias que integran en la comprensión lectora tanto los conocimientos construidos a partir del texto como aquellos otros basados en la memoria a largo plazo. En segundo lugar, el análisis conjunto de las respuestas ofrecidas por los alumnos a preguntas de comprensión inferencial referidas a distintos géneros discursivos reflejó la existencia de diferentes grados de dificultad en la comprensión de un texto.

Por último, las dificultades de comprensión lectora en Costa Rica y España presentaron unos resultados escalados, donde la dificultad asociada a un tipo de comprensión inferencial es similar en ambos países.

Palabras clave: comprensión lectora, género literario, evaluación de la comprensión lectora, estudiantes de educación secundaria

Introduction

Reading comprehension and construction of inferences

Research in the field of reading comprehension has led to a broad consensus regarding the role played by the text's levels (microstructure and macrostructure) in the processes of reading comprehension. Literature reviews conducted by Kendeou (2015) and Perfetti and Adlof (2012) have differentiated between the processes of understanding the microstructure (such as the identification and knowledge of words or the construction of propositions) and the macrostructure (or processes of the discursive level). Van Dijk and Kintsch (1983) outlined in their joint works firstly, and separately, later, this classification of processes that envisaged both the cognitive aspects of reading comprehension and the elements that make up the discourse. As a result, Kintsch (1998) designed the *construction-integration* model of text comprehension, which describes the rules of creation and interconnection of propositions with the text, the activation of prior knowledge and the construction of inferences. The levels of language were tackled by Graesser, Hoffman and Clark (1980), who analysed the time required by the reader to process the text's microstructure and macrostructure. The results of their research showed that the macrostructure required more time and that the purpose of reading influenced the comprehension of the text.

The various theoretical models on reading comprehension have allowed to verify the complexity and the different perspectives that exist in the analysis of the reading processes. Drawing on the *construction-integration model* of Kintsch (1998), the reading comprehension of a text implies the immediate validation of the inferences by the readers.

Received: 04-07-2018 Accepted: 12-04-2019

Readers must monitor the coherence of the text's content, which may include implicit or explicit information. In this way, the analysis of reading comprehension must incorporate the aspects that are related to the construction of the *text-base* as well as the *reference model* (van Dijk & Kintsch, 1983), emphasising the relevance of the interaction between prior knowledge and the text for reading comprehension. The role of the reader, as stressed by Singer (2013, 2014), is to seek the congruence of the reading's meaning, to validate the processes of intentional information retrieval, and to build a coherent reference model. Thus, the reader must validate the elements of the textual microstructure as well as the references to the discursive context, and the knowledge of the world he already owns. Cook and O'Brien's (2014) pointed out the characteristics of this validation. It should be based on the equilibrium and parallel development (activation and integration) of the microstructure and macrostructure processes, and of the knowledge of the world (Cook & O'Brien, 2014). This validation requires the subject's ability to memorise the information, and depends on the information distance and the complexity of access (Singer & Doering, 2014), as well as on the purpose of the reading itself (Smallwood, McSpadden, & Schooler, 2008).

Research on the validation of reading comprehension processes refers to the validation of implicit or explicit notions in the text. When the information is not explicit, the reader must make use of various mechanisms of information retrieval or inferences (Kintsch, 1998). The inferences to reading comprehension are defined as the information generated by the reader during the reading, in order to build a validated reference model (Elbro & Bucj-Iversen, 2013).

Reading comprehension and genres

The construction of a *reference model* requires both the information retrieved in the *text-base* and the integration in the *text-base* of the reader's prior knowledge, related to his experience or his world vision. On the one hand, the reader's information retrieval is partially subject to the recognition of the text's genre. Van Dijk and Kintsch (1983) addressed the predictions and implications that this recognition provides for the retrieval work of the previous information and the construction of the text-base. This also applies to the construction/integration strategies

Received: 04-07-2018 Accepted: 12-04-2019

of the reference model. On the other hand, the development of *genre awareness* in the reader is, according to Johns (2008), the result of the interaction between the social and the cognitive, and between the reader and his construction of flexible discursive prototypes –with specific characteristics for each text level. The reader constructs a mental image from those pragmatic, discursive and linguistic features that determine a genre's textual characteristics, such as a tale, a newspaper column, a scientific article or a sports chronicle (Devitt, 2004).

The relationship between genre's reading comprehension and the processing strategies was initially approached by Fareed (1971) and Graesser, Hoffman and Clark's (1980). The contribution of their research on the genres was the incorporation to the reading comprehension of elements from the *context of culture*, described by Halliday and Hasan (1985). This fact determined, according to Escudero and León (2007) and Zwaan (1994), the incorporation of communication pragmatic elements that have influenced the detection of genre in the reader for an efficient construction of the reference model.

However, we should not overlook the differences developed in the research of reading comprehension between the use of the concept *genre* versus *types of text* (which are, commonly, narrative and expository). Genre, underlines Theresa Lillis (2013), has been developed in a social and situated context, where the reading communication responds to a recurrent situation at a social level and generates a recognisable textual prototype within a discursive community. Research in this field has shown the main role of genre in reading comprehension, due to the influence displayed by the reader's expectations about genre. Zwaan (1994) started from the principle that defines the genres not with a strictly linguistic foundation but based on pragmatic factors. From this perspective, Hyland (2005) and Trabasso and Wiley (2005) have approached the way in which readers adapt their expectations and purposes, both to their knowledge of genre and to the metadiscursive knowledge they possess about it. As a result, the expectations generated by different genres in the reader encourage the development of different reading strategies (Best, Floyd, & McNamara, 2008; Graesser, Singer, & Trabasso, 1994) that influence the decoding and the reader's operational memory. Ignorance of this metadiscursive information, as studied by McNamara, Graesser, and Louwerse (2012), entails great difficulties in the situation model construction, given the impossibility of assembling a coherent mental image.

The reader's previous knowledge related to the text's discursive genre interacts with the information provided by the text-base in the situation model constructing process (Kintsch, 1998). Eason, Goldberg, Young, Geist and Cutting (2012) drew on this conception to affirm that the types of reading comprehension questions are presented differently, according to the genre. Therefore, this research concludes that the characteristics of the texts have an impact on reading comprehension.

The processes by which the reader differentiates the genre during reading have both pragmatic and textual aspects. The first ones refer to the context of the situation in which reading takes place; reading a newspaper column, for instance, discriminates a good part of this context. On the other hand, the text assumes a rhetorical structure that is reflected both in the textual microstructure and in the macrostructure (Biber & Conrad, 2009). The *rhetorical competence* concept allows the readers to detect the text's genre, and allows them to make use of specific strategies that optimise textual comprehension (Sánchez & García, 2009). This rhetorical competence is related to the research of Biber and Conrad (2009) and to the concept of the rhetorical structure of McNamara, Graesser, McCarthy and Cai (2014) regarding the automated analysis of coherence and cohesion of the discourse. Eventually, the reading comprehension of a text makes use of specific strategies in order to build the reference model according to the genre. In such a way, the reader's skills in identifying the genre will depend, largely, on the understanding processes of its macrostructure (Yoo, 2015).

Research on inferential comprehension and genres

Research on reading inferences carried out since the eighties until today has presented two lines of research focused on the models of reading comprehension and genres (McNamara & Magliano, 2009). Since the nineties, several research contributions have focused on the inferential comprehension and the reading comprehension's predictive character. Oakhill and Cain (2012), Kendeou, van den Broek, White and Lynch (2009) and Cromley, Snyder-Hogan and Luciw-Dubas' longitudinal approaches (2010) underline the predictive nature of inferential comprehension in studies conducted in both children and adults.

Received: 04-07-2018 Accepted: 12-04-2019

Research on inferential reading comprehension has been oriented to narrative texts and, to a lesser extent, to expository texts, both in the role played by the genres and the textual cohesion and in the previous knowledge of the world (Best, Ozuru, Floyd, & McNamara, 2006). Some investigations carried out in the eighties (Graesser, Hauft-Smith, Cohen, & Pyles, 1980; Haberlandt & Graesser, 1985; Graesser, Hoffman, & Clark, 1980) already highlighted the differences in reading comprehension between narrative and expository texts. More recently, Graesser, Millis, and Zwaan (1997) related the level of communication of reading with the genre; and Millis and Graesser (1994) carried out a comparative analysis between the types of inferences and the types of discourse. Both perspectives warned of the impossibility of generalising the findings of inferential reading comprehension in narrative texts with respect to other types of texts.

Research on narrative texts has been the most frequent among the studies on inferential comprehension. From the first works on the construction of inferences of Graesser, Singer and Trabasso (1994) to the most recent of Cook and O'Brien (2014) and Gerrig and Wenzel (2015), some publications have addressed reading comprehension of narrative texts, from very different perspectives. In particular, Trabasso and Wiley (2005) state the relevance of the purpose of reading in the understanding of inferences during the comprehension of narrative texts. On the contrary, the most recent works have focused on the analysis of the understanding of expository (McNamara, Louwerse, McCarthy, & Graesser, 2009), scientific (Cromley, Snyder-Hogan, & Luciw-Dubas, 2010; Otero, León, & Graesser, 2002) and argumentative texts (Haria & Midgette, 2014). For example, Kendeou and van den Broek (2007) analyse the effect of prior knowledge and text structure on the reading comprehension cognitive processes in scientific texts.

Comparative studies between the reading of narrative and expository texts show how readers adjust their previous knowledge to reading comprehension. León, Escudero, and van den Broek (2003) replicate the study by Trabasso and Magliano (1996) and incorporate the reading of different genres (narrative, journalistic and expository genre). The results of this research influenced the way in which readers made different types of inferences, according to the type of text they read. In this sense, our article corroborated Kucan and Beck (1996) Goldman, Varma and Coté (1996), Millis and Graesser (1994) and van den Broek, Rohleder

and Narváez's findings (1996), which underscore the different types of inferences made according to the genre.

Our current research

The recent reviews on inferential reading comprehension (Gerrig & Wenzel, 2015; Lorch, 2015) show the interest developed by the textual macropropositions of narrative and expository sequences, but it is far from reading in a natural context (Steen, 1999). However, there are very few investigations that analyse the concept of genre (Adam, 2001) in relation to the construction of inferences in reading. The research is aimed at the analysis of the behaviour of different inferences in reading (van den Broek, Fletcher & Risden, 1993, van den Broek, Beker, & Oudega, 2015) specific genres (journalistic column, sports chronicle and Tale). In this sense, this article responds to a working line that Graesser, Gernsbacher and Goldman (1997) had pointed out as a line of research within the framework of reading comprehension. In this direction, our research questions are the following:

- What are the differences in the construction of the types of inferences in the reading of different genres (journalistic column, sports chronicle and Tale) in the students, according to their age, their sex and to the tenure of the educational establishment?
- What are the differences in the inferential understanding between secondary education students in Costa Rica and Spain?

Methodological framework

Sample

The sample included 3,199 secondary education students (12, 14 and 16 years old) from two countries, Spain and Costa Rica, that have been chosen intentionally. In Spain, a total of 2,490 subjects enrolled in 17 public and private educational centres of the province of Seville took part in the sample. In Costa Rica, the participation of a total of 5 schools (public, private and semi-private educational establishments in rural

Received: 04-07-2018 Accepted: 12-04-2019

and urban contexts) was considered, with a total of 709 students from the provinces of San José and Cartago. As we may observe in Table 1, the distribution of the sample according to sex, age and tenure of the educational centre was balanced.

TABLE I. Description of the sample

Study variables		Frequency	Percentage
Country	Spain	2,490	77.8
	Costa Rica	709	22.2
Gender	Female	1,562	48.8
	Male	1,630	51.0
Age	12 years old	1,173	36.7
	14 years old	1,065	33.3
	16 years old	961	30.0
Ownership of Secondary School	Public	1,798	56.2
	Private	1,401	43.8
Total sample		3,199	100.0

Procedure for collecting information

In order to respond to the research questions, set out in this work, the inferential reading comprehension test (TEst de COmprensión LEctora INferencial – TECOLEIN) was designed and validated. This instrument is composed of three texts and a set of comprehension questions. The texts integrated some of the secondary education curriculum's genres of Costa Rica and Spain. The test's first text is an Article entitled "Civismo" (Montero, 2012), which is followed by a sports chronicle entitled "Berdych elimina a Nadal" (Mateo, 2015). Both texts were taken from the newspaper *El País*. The third text consisted of a Costa Rican Tale entitled "El bongo" (Salazar Herrera, 1947). The three texts were selected for being representative of their genres and for possessing enough inferences in order to elaborate the reading comprehension questions of TECOLEIN.

The texts and the questions contained in TECOLEIN were drawn up by three researchers, two from Spain and another from Costa Rica.

The elaboration process removed the questions that at least one of the researchers did not consider appropriate, due to its content or format. Likewise, the different versions of the test were reviewed by four experts in reading comprehension and measurement. Two pilot tests were carried out both in Spain and in Costa Rica until the final test was set; in none of them were the texts or the questions repeated. The first pilot test, which differentiated between 4 modalities (printed/digital, fragmented/complete), was applied to 1,266 students from Spain and Costa Rica. The results obtained did not allow to establish statistically significant differences between the printed and digital modalities. This is why the digital modality was ruled out, given the lack of computing resources observed in some centres of the final projected sample. The fragmented text mode was chosen since it allowed to evaluate a greater number of types of inferences. The items that presented low levels of discrimination were reformulated and then reapplied to a sample of 709 students.

Table 2 shows the matrix of TECOLEIN specifications. Causal reverse inferences are composed of reverse connective causal inferences (reactivate information from short-term memory of the text) and elaborative inferences (retrieve information from long-term memory). The forward elaborative inferences relate information that is not relevant to the logical structure of the discourse, but that affects the coherence of the text. Ultimately, the thematic inference allows the construction of the global coherence of reading, based on the text's main idea and on the author's intention (van den Broek, 1994; van den Broek, Beker, & Oudega, 2015).

TABLE II. Table of specifications for the description of the instrument

Inferences / Genre	Connective causal inferences	Backward causal inferences	Forward elaborative causal inferences	Thematic inference	Total
Article	2, 3, 5, 7, 10	1, 6, 8, 11	-	4, 9, 12	12
Sports chronicle	4, 6, 7, 9	1, 2, 3, 10	11	5, 8, 12	12
Tale	5, 8, 11, 12	1, 2, 4, 7	3, 10	6, 9, 13	13
Total	13	12	3	9	37

Received: 04-07-2018 Accepted: 12-04-2019

In its final version, TECOLEIN was completed on paper by students, in their classrooms, in a maximum period of one hour. The texts that compose TECOLEIN are fragmented into three parts, each one of them with 3 to 6 comprehension questions. The questions are multiple choice questions with four response modes and only one is correct. For example, question 12 (thematic inference), which refers to the first part of the text "Berdych elimina a Nadal" (sports chronicle), asks:

12. The main idea of the fragment should include both the subject of the previously read text and the author's intention when writing it. In your opinion, what is the main idea of the previous fragment?

a. Nadal's defeat is received with surprise by the public.
b. Nadal's physical difficulties prevented him from winning the game.
c. Nadal's physical problems offer an opportunity to Berdych.
d. Berdych's game takes advantage of Nadal's limitations.

The students could read the text several times but, once the reading was concluded, they were not allowed to "go back". The answers were evaluated in terms of success and error and the response time was not measured. The scores were obtained considering the type of inference and text, by summing the correct answers. Most of the TECOLEIN items have medium difficulty indexes (between .40 and .60), being .27 and .93 the extreme values. The most difficult items correspond to the "Article" genre (average difficulty of .53) and to the "thematic" inference (average difficulty of .51). TECOLEIN allows to discriminate between different levels of inferential reading comprehension. The items that best discriminate are those referred to the "chronic sports" genre (.67) and the "thematic" inference (.58). The instrument is internally consistent, with a Cronbach's alpha of .77. TECOLEIN considers that the inferential reading comprehension construct is defined according to the genre and according to the type of text that is presented to the reader. This definition can be considered sound according to the experts' assessment and to the empirical verification, which is carried out using the Confirmatory Factor Analysis (CFA) method, with the Maximum Likelihood adjustment function. The Chi-square absolute fit indices obtained was 171,566, for 31 gl with a significance of .001, and with values RMSEA=.038, RMR=.033, GFI=.990 y AGFI=.979.

Received: 04-07-2018 Accepted: 12-04-2019

Data analysis

In order to give answers to the objectives of this work, we used different data analysis techniques. The *t* test for paired samples allowed us to identify what type of text and inference generate the greatest difficulty in the students. To do this, we calculated the average score expected for each of the items (average of the maximum score in a text or inference considering the number of affected items) and then we compared it with the average score obtained by the students in said test. We also carried out an ANOVA test of repeated measures and a *t* test of paired samples in order to contrast the existence of differences in inferential comprehension according to the genres.

The two-stage cluster analysis was used to group the answers of the students according to personal (sex and age) and contextual variables (country and tenure of the centre). The measure of distance used has been Maximum Likelihood and the number of clusters has been specified according to the number of groups of each variable, using the Schwarz Bayesian Information Criterion (BIC).

Results

Responses of the students to TECOLEIN

The students' answer to TECOLEIN offer different reading comprehension results according to the types of texts and inferences. The *difference test analysis with the midpoint* has allowed to identify in what type of text and inference the answers are higher or lower than what would be expected in each case. The *midpoint* column represents the expected average value and the *average difference* column collects the differences in punctuation between the expected average values and those obtained by the students. When this latter value is negative, it indicates a lower performance of the students than expected for a given genre and inference. This is the case of the results obtained for the *Article* (average dif.:1.18) and *subject-text inference* (average dif.:0.49); in the remaining types of texts and inferences, the mean value expected is greater than that obtained (Table 3).

Received: 04-07-2018 Accepted: 12-04-2019

TABLE III. Difference test analysis with the midpoint according to the genre and the type of inference

Genres/Inferences	Item	Mid-point	Descriptive analysis			Sig.
			Average difference	Std.	t	
Article	12	6	*-1.18*	1.71	-30.682	.0001
Sports chronicle	13	6.5	3.91	2.58	67.228	.0001
Tale	13	6.5	4.22	2.92	64.175	.0001
Connective causal inferences	13	6.5	3.29	2.22	65.947	.0001
Backward causal inferences	12	6	1.16	2.13	24.221	.0001
Forward elaborative causal inferences	3	1.50	0.47	0.93	22.297	.0001
Thematic inference	9	4.5	*-0.49*	1.53	-14.315	.0001

An analysis of the answers in relation to the text *Article* (average dif.= -1.18) reveals that the students have had more difficulty answering the questions that needed a reverse causal inference. The lowest expected average value is that obtained from the joint scores between the Article and the reverse causal inference (average dif. = -0.37) (Table 4). These results would indicate that not all inferences present the same level of difficulty before the same genre.

TABLE IV. Difference test analysis with the midpoint according to the genre and the type of inference

Genres /Inferences	Item	Mid-point	Descriptive analysis			Sig.
			Average difference	Std.	t	
Article-causal connective inferences	5	2.5	0.81	1.17	30.910	.0001
Article- Backward causal inferences	4	2	*-0.37*	1.00	-16.201	.0001
Article- Thematic inference	3	1.5	0.29	0.89	14.426	.0001

The joint analysis of the *thematic inference of the text* (average dif. = -0.49) and of the different genres shows that students have a greater difficulty in making this type of inference. In the three genres we analysed, the thematic inference always presents values that are lower than the expected average (Table 5).

TABLE V. Difference test analysis with the midpoint according to the genre and the type of inference

Genres/Inferences	Item	Midpoint	Descriptive analysis			Sig.
			Average difference	Std.	t	
Opinion C.-Thematic inference	3	1.5	*-0.29*	0.88517	14.426	.0001
Sports Chr. - Thematic inference	3	1.5	*-0.22*	0.86339	-11.480	.0001
S. Story- Thematic inference	3	1.5	*-0.56*	0.76712	-32.251	.0001

The repeated measures analysis indicates that the difficulty and the way in which a certain type of inference is read is influenced by the genre. Statistically significant differences were found in each one of the types of inferences, according to the genres. The *Cohen's partial Eta squared and d* values indicate a moderate overall effect of the genre on the type of inference, with the exception of the connective causal inference (Ruscio & Mullen, 2012). The differences observed in the inferential comprehension according to the genre present the same pattern in Costa Rica and Spain. The highest average scores in Spanish students in a genre are also higher in students from Costa Rica (Table 6).

Received: 04-07-2018 Accepted: 12-04-2019

TABLE VI. Analysis of the types of inferences according to the genres

Types of inferences/Genres		F		gl		Sig.		Partial Eta squared value		$\bar{x}$		Std.	
		ES	CR	ES	CR	ES	CR	ES	CR	ES	CR	ES	CR
Connective causal inferences	Opinion C.	198.68	13.08	2.469	707	0.000	0,000	0.139	0.036	3.39	3.19	1.17	1.18
	Sports Chr.									3.61	2.95	0.70	1.06
	S. Story									3.21	2.92	0.98	1.18
Backward causal inferences	Opinion C.	1362.24	218.50	2.465	707	0.000	0,000	0.525	0.382	1.70	1.44	1.04	0.91
	Sports Chr.									2.92	2.32	1.01	1.08
	S. Story									2.95	2.52	1.01	1.13
Forward elaborative causal inferences	Sports Chr.	-49.14 (t)	27.41	2.477	708	0.000	0,000	0.540 *r* based on Cohen's *d*	0.550 *r* based on Cohen's *d*	0.68	0.45	0.47	0.50
	S. Story									1.42	1.28	0.67	0.74
Thematic inferences	Opinion C.	825.58	164.11	2.468	707	0.000	0,000	0.401	0.317	1.89	1.66	0.85	0.92
	Sports Chr.									1.73	1.05	5.63	0.83
	S. Story									0.96	0.93	0.78	0.74

Grouping of the answers of the students

The two-stage cluster analysis has revealed the student's patterns of responses, based on personal and contextual variables (country, sex, age and tenure). The two-stage cluster analysis has generated two clusters based on the Schwarz Bayesian Information Criterion (BIC= 4570.81) and the Maximum Likelihood method (distance measurement ratio=1.262). In cluster A there are 1,223 people (38.2%); in cluster B there are 1,786 people (55.8%). Results show that there are statistically significant differences between clusters A and B, in all variables except *gender*. Cluster A is composed only of Spanish students (50.1%), who study in privately owned centres (89.9%) and of students of all ages (12, 14 and 16 years). Instead, Cluster B is made up of 96.5% of Costa Rican students and 49.3% of Spanish students. They all study in public centres.

TABLE VII. Comparison of demographic data

Variables	Total (n=3,199)		Cluster				Chi-square	gl	Sig.
			A (n=1,223)		B (n=1,786)				
	n	(%)	*n*	(%)	*n*	(%)			
Country									
Spain	2,490	77.8	1,223	50.1	1,204	49.3	521.26	2	<.001
Costa Rica	709	22.2	0	0.0	582	96.5			
Gender									
Male	1,562	48.8	624	40.2	906	58.4	1.185	2	.553
Female	1,630	51.0	599	40.1	880	58.9			
Age									
12 years old	1,173	36.7	500	44.9	589	52.9	54.90	4	<.001
14 years old	1,065	33.3	427	41.7	587	57.4			
16 years old	961	30.0	296	32.6	610	67.2			
Tenure									
Public	1,798	56.2	0	0.0	1,662	98.8	2543.23	2	<.001
Private	1,401	43.8	1,223	89.9	124	9.1			

Schwarz Bayesian Information Criterion (BIC) = 4570,811; BIC change = -7,564; Distance measurement ratio = 1,262

The scores obtained by the students in the different types of texts and inferences present statistically significant differences (p≤.002). In all cases, the differences are in favour of the students grouped in Cluster A. Their scores are higher and less dispersed with respect to the cluster B average. The scores reached by the students are higher in the *sports chronicle* ($\bar{x}$=8.87) and in the *Tale* ($\bar{x}$=8.63), and lower in the Article ($\bar{x}$=6.50). In relation to the types of inferences, students reach higher scores when they must make *connective causal inferences* ($\bar{x}$=10.40), and lower, in the forward elaborative causal inference ($\bar{x}$=2.18).

Received: 04-07-2018 Accepted: 12-04-2019

TABLE VIII. Comparison according to genres and types of inferences

TECOLEIN	Cluster					T-test	gl	Sig.
	Nº items	A (n=1223)		B (n=1786)				
		$\bar{x}$	Sx	$\bar{x}$	Sx			
Genre								
Article	12	6.50	1.94	6.06	1.93	6.140	3.007	<.001
Sports chronicle	12	8.87	1.90	7.97	2.27	11.759	2.894	<.001*
Tale	13	8.63	2.11	8.36	2.42	3.279	2.837	<.001*
Type of inference								
Connective causal inference	13	10.40	1.87	9.87	2.14	7.243	2.833	<.001*
Backward causal inferences	12	7.76	1.99	7.11	2.11	8.524	3.007	<.001
Forward elaborative causal inference	3	2.18	0.84	1.98	0.91	5.969	3.007	<.001
Thematic inference	9	4.28	1.43	4.10	1.50	3.162	3.007	.002

*Sig. <,001 Levene test for equality of variances

Clusters A and B show that the scores of students in all genres correlate significantly (p <.001) with all types of inferences. The different types of inferences made by students change significantly and with different intensity, depending on the texts they read. For example, cluster A has a correlation of .279 between the scores in the *forward elaborative causal inference* and the *Article* scores. On the other hand, the correlation of this inference with the scores in the sports chronicle is of .513.

The results of this investigation show a statistically significant correlation (p <0.001) between the three genres and the four types of inferences measured with the total score corrected in TECOLEIN. This score reflects the variability observed in the responses of the students to the questions which are related to different genres that require different types of inferences. Data show that the highest correlations correspond to the *Tale* and to the *connective causal* and *Backward causal inferences inferences*. Instead, the lowest correlations are found in the *Article* and in the *thematic inferences* of the text and in *forward elaborative causal inference*. This same pattern of results is observed in the correlations between the genres and the types of inferences and the total corrected TECOLEIN score.

TABLE IX. Correlation matrix

	Cluster	Article	Sports Chr.	Tale	Connective causal inference	Backward causal inferences	Forward elaborative causal inference	Thematic inference	TECOL.
Article		I							
Sports chronicle	A *B*	.310 *.328*	I						
Tale	A *B*	.333 *.317*	.349 *.440*	I					
Connective causal inference	A *B*	.606 *.581*	.463 *.557*	.630 *.657*	I				
Backward causal inferences inference	A *B*	.641 *.594*	.593 *.654*	.567 *.617*	.479 *.493*	I			
Forward elaborative causal inference	A *B*	.279 *.308*	.513 *.561*	.572 *.615*	.380 *.453*	.365 *.435*	I		
Thematic inference	A *B*	.420 *.441*	.515 *.548*	.444 *.497*	.263 *.319*	.265 *.357*	.311 *.343*	I	
TECOLEIN [1]	A *B*	.649 *.607*	.646 *.711*	.693 *.729*	.717 *.743*	.740 *.761*	.571 *.626*	.538 *.588*	I

[1] Score of the item corrected in TECOLEIN

Discussion

The scores obtained in TECOLEIN make it possible to measure the inferential comprehension of the students before questions related to genres demanding different types of inferences (McKoon & Ratcliff, 2015). Firstly, this measurement allows us to observe a graduation of inferential reading comprehension. The scores of 12, 14 and 16-year-old students show statistically significant differences ($p < 0.001$) according to the genres and the types of inferences. In the same line proposed by Ozuru, Rowe, O'Reilly and McNamara (2008), these differences are an indicator of the level of reading comprehension developed by students during secondary education. These data corroborate how the reading skills related to the macrostructure are developed progressively from 12 to 16 years for a wide number of genres. The processes that monitor

Received: 04-07-2018 Accepted: 12-04-2019

reading comprehension, working memory and, especially, the generation of inferences are significant indicators of reading development (Kendeou, 2015). In particular, the generation of inferences during the reading of a text becomes one of the instrumental skills that students develop during secondary education. The progressive improvement observed in the results of TECOLEIN shows how students have developed strategies that integrate both the knowledge constructed on the basis of the text and the one based on long-term memory in the reading comprehension (Cook & O'Brien, 2014; Singer & Richards, 2005; van den Broek, Rapp, & Kendeou, 2005).

Secondly, the joint analysis of the answers provided by the students to questions of inferential comprehension, which are referred to different genres, reflects the existence of different degrees of difficulty in the comprehension of a text. Research by Trabasso and Magliano (1996) and León, Escudero and van den Broek (2003) already showed that readers construct different inferences according to the type of text. Our research goes further, showing how the same type of inference has a different degree of difficulty according to the genre in which it is read. This support the thesis of McNamara, Graesser and Louwerse (2012), who underlined the relationship between reading comprehension and textual cohesion in different discourses. The data analysed in our work challenge the idea that the difficulty in reading inferential comprehension is associated either with the genre or with the type of inference (Lorch, 2015). The combination of metadiscursive knowledge about the genre and the types of inference in this genre is what actually determines the degree of difficulty in understanding a text. In other words, the difficulty to understand a text depends on the reader's prior knowledge (long-term memory), on its ability to predict the genre and, finally, on the validation process that the reader performs through the integration of all this information in the construction of the reference model (Singer, 2015).

An educational application of this research is the incorporation into the classroom of a *pedagogy based on genres* of the Sydney School (Rose and Martin, 2012). Our research emphasises the need to incorporate reading comprehension practices from real and everyday genres. These reading practices would allow to overcome the limitation of the construction of the meaning of a type of inference in different genres (Eason et al., 2012). On the one hand, writing and reading in all the

disciplines of the curriculum would acquire a fundamental role for the development of reading comprehension in secondary education; on the other hand, the creation of an intervention measure for the improvement of reading comprehension must be based not only on the development of general strategies and reading techniques, but on the adaptation of these techniques to a specific reading situation (Baker, Gersten, & Grossen, 2002; Best et al., 2008), to the integration of information, the validation process according to the purposes of reading and to the characteristics of the discursive genre (Kintsch, 1998; McNamara, Kintsch, Songer, & Kintsch, 1996).

Thirdly, the scores obtained in TECOLEIN present statistically significant differences in the inferential comprehension of students from Spain and from Costa Rica. These differences in favour of the Spanish sample affect all types of genres and inferences we evaluated. Nevertheless, the different levels of understanding between the students of both countries are scaled. The genres that present higher scores in Spain tend to be the highest in Costa Rica and vice versa. Similarly, the difficulty associated with a type of inferential understanding is similar in both countries. The results obtained in TECOLEIN are coherent with the score obtained by both countries in PISA 2015 (OECD, 2016).

The differences found in TECOLEIN highlight the educational challenges faced by Costa Rica and Spain. Inferential reading comprehension has become one of the clearest indicators of the reading level. TECOLEIN emphasises the need to incorporate a reader process which is incardinated in the current social context. In this way, the curriculum should address reading from a plural and heterogeneous perspective, open to the development of inferential comprehension in new forms of reading in Web 2.0, in television or in orality (Magliano, Loschky, Clinton, & Larson, 2013; Kendeou, 2015). The inferential reading comprehension, in this sense, will depend on the progressive development of the metadiscursive abilities in the reader, abilities that allows him to approach the reading learning throughout all his life.

In conclusion, this article highlights the importance of inferences in the process of reading comprehension in secondary education. The development of reading comprehension in this educational stage presents a fundamental role centred on the generation of inferences. The secondary education curriculum also raises a great diversity of school genres in its disciplines, creating new strategies of reading

comprehension in students. Our research has shown the progressive improvement of TECOLEIN results, as students develop new reading comprehension strategies throughout secondary education. Likewise, this article confirms how the difficulties of inferential comprehension are analogous in different countries, such as Costa Rica and Spain. The differences found between the two countries highlight precisely how students face similar difficulties in reading different genres. Finally, this research opens new ways to optimise learning at this stage, in which metadiscursive knowledge and inferential comprehension strategies in the disciplines could substantially improve school performance. The improvement of inferential comprehension will depend on the strategies that students put into practice in their reading and on the ability to apply these strategies to different genres.

Limitations

This article poses some limitations related to the research process on inferential reading comprehension. First of all, the research was not developed in a context of natural reading, where the reader assumes a leading role in his own reading, by choosing the document and the purpose of reading (Lorch, 2015). Nevertheless, TECOLEIN texts were complete and its extension was not modified for the test. Student errors were not reduced by the use of abbreviated texts.

Secondly, the differences found between the inferential reading comprehension of the genres studied in TECOLEIN (Article, sports chronicle and Tale) should not be compared with the previous studies on reading comprehension based on the types of texts. Studies that focus on the differences in reading comprehension of genres allow to incorporate into the research elements present in natural reading, such as prior knowledge, text complexity and metadiscursive skills that the reader uses. This research, however, has not been able to determine which aspects of the information integration process, the improvement of the intentional focus and the development of the executive functions have determined the results of TECOLEIN (Diamond, 2013; Liu, Reichle, & Gao, 2013).

Lastly, our research has not analysed the motivations of the students in relation to the texts or their previous knowledge (Halldorson &

Received: 04-07-2018 Accepted: 12-04-2019

Singer, 2002, McKoon & Ratcliff, 1988, Singer, 2015). This aspect could also influence the achievement of reading comprehension of a text. In subsequent studies, the TECOLEIN test should incorporate some questions that would allow to relate the motivation of students in relation to the efficiency of inferential reading comprehension.

References

Adam, J.-M. (2001). Types de textes ou genres de discours? Comment classer les textes qui *dissent de et comment faire? Langages, 141*, 10-27.

Baker, S., Gernsten, R., & Grossen, B. (2002). Interventions for students with reading comprehension problems. In M.R. Shinn, H.M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Prevention and remedial approaches* (pp. 731-754). Bethesda, MD: National Association of School Psychologists.

Best, R.M., Floyd, R.G., & McNamara, D.S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology, 29*, 137-164. Doi: 10.1080/02702710801963951

Best, R.M., Ozuzu, Y., Floyd, R.G. & McNamara, D.S. (2006). Children's text comprehension: effects of genre, knowledge, and text cohesion. En S.A. Barab, K.E. Hay & D.T. Hickey (Eds.), *Proceedings of the seventh international conference of the learning sciences* (pp. 37-42). Mahwah, NJ: Erlbaum.

Biber, D. & Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.

Cook, A.E. & O'Brien, E.J. (2014). Knowledge activation, integration, and validation during narrative text comprehension. *Discourse Processes, 51*, 26-49.

Cromley, J.G., Snyder-Hogan, L.E., & Luciw-Dubas, U.A. (2010). Reading comprehension of scientific text: a domain-specific test of the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 102*(3), 687-700.

Devitt, A.J. (2004). *Writing genre*. Carbondale: Southern Illinois University.

Received: 04-07-2018 Accepted: 12-04-2019

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135-168. doi: 10.1146/annurev-psych-113011-143750

Eason, S.H., Goldberg, L.F., Young, K.M., Geist, M.C., & Cutting, L.E. (2012). Reader-Test interactions: How differential text and question types influence cognitive skills needed for reading comprehension. *Journal of Educational Psychology, 104*(3), 515-528.

Escudero, I. & León, J.A. (2007). Procesos inferenciales en la comprensión del discurso escrito. Influencia de la estructura del texto en los procesos de comprensión. *Revista Signos, 40*(64), 311-336.

Elbro, C. & Buch-Iversen, I. (2013). Activation of background knowledge for inference making: effects on reading comprehension. *Scientific Studies of Reading, 17*, 435-452.

Fareed, A.A. (1971). Interpretative responses in reading history and biology: an exploratory study. *Reading Research Quarterly, 6*(4), 493-532.

Gerrig, R.J. & Wenzel, W.G. (2015). The role of inferences in narrative experiences. In E.J. O'Brien, A.E. Cook, & R.F. Lorch, Jr. (Eds.), *Inferences during reading* (pp. 362-385). Cambridge: Cambridge University Press.

Goldman, S.R., Varma, S. & Coté, N. (1996). Extending capacity-constrained construction integration: Toward "smarter" and flexible models of text comprehension. En B.K. Britton & A.C. Graesser (Eds.), *Models of understanding text* (pp. 73-113). Hillsdale, NJ: Erlbaum.

Graesser, A.C., Gernsbacher, M.A., & Goldman, S.R. (1997). Cognition. En T. A. van Dijk (Ed.), *Discourse: A multidisciplinary introduction* (pp. 292-319). London: Sage.

Graesser, A.C., Hauft-Smith, K., Cohen, A.D., & Pyles, L.D. (1980). Advanced outlines, familiarity, and text genre on retention of prose. *Journal of Experimental Education, 48*(4), 281-290.

Graesser, A.C., Hoffman, N.L., Clark, L.F. (1980). Structural components of reading time. *Journal of Verbal Learning and Verbal Behavior, 19*, 135-151.

Graesser, A.C., Millis, K.K., & Zwaan, R.A. (1997). Discourse comprehension. *Annual Review of Psychology, 48*, 163-189.

Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*, 371-395.

Halliday, M.A.K. & Hasan, R. (1985). *Language, context, and text: aspects of language in a social-semiotic perspective*. Burwood, Victoria: Deakin University.

Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General, 114*(3), 357-374. doi: 10.1037/0096-3445.114.3.357

Halldorson, M. & Singer, M. (2002). Inference processes: Integrating relevant knowledge and text information. *Discourse Processes, 34*(2), 145-161. doi: 10.1207/S15326950DP3402_2

Haria, P.D. & Midgette, E. (2014). A genre-specific reading comprehension strategy to enhance struggling fifth-grade readers' ability to critically analyze argumentative text. *Reading & Writing Quarterly, 30*, 297-327. Doi: 10.1080/10573569.2013.818908

Hyland, K. (2005). *Metadiscourse. Exploring interaction in writing*. London: Continuum.

John, A.M. (2008). Genre awareness for the novice academic student: an ongoing quest. *Language Teaching, 41*(2), 237-252. DOI: 0.1017/S0261444807004892

Kendeou, P. (2015). A general inference skill. En E.J. O'Brien, A.E. Cook, & R.F. Lorch, Jr. (Eds.), *Inferences during reading* (pp. 160-181). Cambridge: Cambridge University Press.

Kendeou, P. & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition, 35*(7), 1567-1577.

Kendeou, P., van den Broek, P., White, M.J., & Lynch, J.S. (2009). Predicting reading comprehension in early elementary school: the independent contributions of oral language and decoding skills. *Journal of Educational Psychology, 101* (4), 765-778.

Kintsch, W. (1998). *Comprehension. A paradigm for cognition*. Cambridge: Cambridge University Press.

Kucan, L. & Beck, I.L. (1996). Four fourth graders thinking aloud: an investigation of genre effects. *Journal of Literacy Research, 28*(2), 259-287.

León, J.A., Escudero, I., & van den Broek, P. (2003). La influencia del género del texto en el establecimiento de inferencias elaborativas. En J.A. León (Coor.), *Conocimiento y discurso. Claves para inferir y comprender* (pp. 153-170). Madrid: Pirámide.

Lillis, T. (2013). *The sociolinguistics of writing*. Edinburgh: Edinburgh University Press.

Liu, Y.P., Reichle, E.D. & Gao, D.G. (2013). Using reinforcement learning to examine dynamic attention allocation during Reading. *Cognitive Science, 37*(8), 1507-1540. doi: 10.1111/cogs.12027

Lorch, R.F., Jr. (2015). What about expository text? En E.J. O'Brien, A.E. Cook, & R.F. Lorch, Jr. (Eds.), *Inferences during reading* (pp. 348-361). Cambridge: Cambridge University Press.

Magliano, J.P., Loschky, L.C., Clinton, J., & Larson, A.M. (2013). Is reading the same as viewing? An exploration of the similarities and differences. Between processing text- and visually based narratives. En B. Miller, L. Cutting, & P. McCardle (Eds.), *Unraveling the Behavioral, Neurobiological, and Genetic Components of Reading Comprehension* (pp. 78-90). Baltimore, MD: Brookes Publishing.

Mateo, J.J. (27 de enero de 2015). Berdych elimina a Nadal. *El País*. Recuperado de https://elpais.com/deportes/2015/01/27/actualidad/1422337869_245842.html

McKoon, G. & Ratcliff, R. (1988). Contextually relevant aspects of meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(2), 331-343.

McKoon, G. & Ratcliff, R. (2015). Cognitive theories in discourse-processing research. En E.J. O'Brien, A.E. Cook, & R.F. Lorch, Jr. (Eds.), *Inferences during reading* (pp. 42-67). Cambridge: Cambridge University Press.

McNamara, D.S., Graesser, A.C., & Louwerse, M.M. (2012). Sources of text difficulty: across genres and grades. In J. Sabatini, E. Albro, & T. O'Reilly (Eds.), *Measuring up. Advances in how we assess reading ability* (pp. 89-118). Lanham: Rowman & Littlefield Education.

McNamara, D.S., Kintsch, E., Songer, N.B. & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*(1), 1-43.

McNamara, D.S. & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of learning and motivation, 51*, 297-384. doi: 10.1016/S0079-7421(09)51009-2

McNamara, D.S., Graesser, A.C., McCarthy, P.M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge: Cambridge University Press.

Millis, K.K. y Graesser, A.C. (1994). The time-course of constructing knowledge-based inferences for scientific texts. *Journal of Memory and Language, 33*, 583-599.

Montero, R. (21 de febrero de 2012). Civismo. *El País*. Recuperado de https://elpais.com/elpais/2012/02/20/opinion/1329739526_020260.html

Oakhill, J.V. & Cain, K. (2012). The precursors of Reading ability in Young readers: evidence from a four-year longitudinal study. *Scientific Studies of Reading, 16*(2), 91-121.

OECD (2016). *PISA 2015 Results (Volume I): Excellence and equity in Education*. Paris: OECD Publishing. doi: 10.1787/19963777

Otero, J., León, J.A. & Graesser, A.C. (Eds.) (2002). *The psychology of science text comprehension*. Mahwah, NJ: Routledge.

Ozuru, Y., Rowe, M., O'Reilly, T., & McNamara, D.S. (2008). Where's the difficulty in standardized reading tests: The passage or the question? *Behavior Research Methods, 40*(4), 1001-1015. Doi: 10.3758/BRM.40.4.1001

Perfetti, C.A., & Adlof, S.M. (2012). Reading comprehension: a conceptual framework from word meaning to text meaning. In J.P. Sabatini, E.R. Albro, & T.O'Reilly (Eds.), *Meaning up: Advances in how to assess reading ability* (pp. 3-20). Lanham, JH: Rowman & Littlefield Education.

Rose, D. & Martin, J.R. (2012). Reading to learn. In D. Rose & J.R. Martin, *Learning to write, reading to learn. Genre, knowledge and pedagogy in the Sydney School* (pp. 133-234). Briston (CT)-South Yorkshire: Equinox.

Ruscio, J., & Mullen, T. (2012). Confidence Intervals for the Probability of Superiority Effect Size Measure and the Area Under a Receiver Operating Characteristic Curve. *Multivariate Behavioral Research*, 47(2), 201–223.

Salazar Herrera, C. (1990). El bongo. In C. Salazar Herrera, *Cuentos de angustias y paisajes* (pp. 39-44). San José: Editorial El bongo.

Sánchez, E. & García, J.R. (2009). The relation of knowledge of textual integration devices to expository text comprehension under different assessment conditions. *Reading & Writing, 22*, 1081-1108. Doi: 10.1007/s11145-008-9145-7

Singer, M. (2013). Validation in reading comprehension. *Current Directions in Psychological Science, 22*, 361-366.

Singer, M. (2015). Validation of text and discourse inferences — and explicit content. In E.J. O'Brien, A.E. Cook, & R.F. Lorch, Jr. (Eds.), *Inferences during reading* (pp. 68-93). Cambridge: Cambridge University Press.

Singer, M. & Doering, J.C. (2014). Exploring individual differences in language validation. *Discourse Processes, 51*, 167-188.

Singer, M. & Richards, E. (2005). Representing complex narrative goal structures: competing memory-based and situational influences. *Discourse Processes, 39*(2&3), 189-204.

Smallwood, J., McSpadden, M., & Schooler, J.W. (2008). When attention matters: the curious incident of the wandering mind. *Memory & Cognition, 36*, 1144-1150.

Steen, G. (1999). Genres of discourse and the definition of literature. *Discourse Processes, 28*(2), 109-120. Doi: 10.1080/01638539909545075

Trabasso, T. & Magliano, J.P. (1996). Conscious understanding during comprehension. *Discourse Processes, 21*, 255-287.

Trabasso, T. & Wiley, J. (2005). Goal plans of action and inferences during comprehension of narratives. *Discourse Processes, 39*(2&3), 129-164.

Van den Broek, P. (1994). Comprehension and memory of narrative texts. In M.A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 539-588). San Diego: Academic Press.

Van den Broek, P., Fletcher, C.R, & Risden, K. (1993). Investigations of inferential processes in reading: a theoretical and methodological integration. *Discourse Processes, 16*, 169-180.

Van den Broek, P., Beker, K., & Oudega, M. (2015). Inference generation in text comprehension: automatic and strategic processes in the construction of a mental representation. In E.J. O'Brien, A.E. Cook, & R.F. Lorch, Jr. (Eds.), *Inferences during reading* (pp. 94-121). Cambridge: Cambridge University Press

Van den Broek, P., Rapp, D.N., & Kendeou, P. (2005). Integrating memory-based and constructionanist processes in accounts of reading comprehension. *Discourse Processes, 39*(2&3), 299-316.

Yoo, M.S. (2015). The influence of genre understanding on strategy use and comprehension. *Journal of Adolescent & Adult Literacy, 59*(1), 83-93. Doi: 10.1002/jaal.432

Zwaan, R.A. (1994). Effect of genre expectations on text comprehension. *Journal of Experimental Psychology, 20*(4), 920-933.

Contact address: Fernando Guzmán-Simón. Universidad de Sevilla, Facultad de Ciencias de la Educación, Departamento de Didáctica de la Lengua y la Literatura. C/ Pirotecnia, s/n, 41013, Sevilla. E-mail: fernandoguzman@us.es

Critical analysis of university student satisfaction surveys[1]

Análisis crítico de las encuestas universitarias de satisfacción docente

DOI: 10.4438/1988-592X-RE-2019-385-418

Vicente Llorent-Bedmar
Verónica C. Cobano-Delgado Palma
Universidad de Sevilla

Abstract

At present, most universities in developed countries carry out surveys of university students' satisfaction with the teaching received. There is widespread debate on the relevance and true purpose of this assessment.

In order to contribute to the creation of a truly useful assessment tool with the potential to help improve university teaching, we have analysed the satisfaction surveys used in Spanish public universities using qualitative research based on content analysis. The corpus of analysis was Spanish public universities (N=47). We coded 711 items in all according to measurements and themes, depending on the aspect assessed in each item, following an inductive process based on a semantic criterion. Inter-coder reliability was calculated using Krippendorff's alpha, SPSS statistical software and the Hayes macro for nominal variables. Human coding was carried out for content analysis using atlas.ti qualitative software.

From the most important results it was concluded that there are aspects which do not depend on the task of the teacher directly or indirectly but are assessed by calculating the mean with the rest of variables; others that are not the direct responsibility of the teachers; some aspects that are important to teaching but are not assessed; and in most cases, the predominant formulation

(1) This paper has been funded by Grupo de Investigación de Educación Comparada de Sevilla, GIECSE (Code HUM-486, Plan Andaluz de Investigación, Desarrollo e Innovación of the Junta de Andalucía) in the framework of the project `Ayuda a Grupos 2017´ of the Dirección General de Investigación y Transferencia del Conocimiento (Junta de Andalucía).

of closed items which limit valuable information potentially useful for improving teachers' activity. In conclusion, there is a need to redesign these tools, taking into account the perspective of students as well as that of teachers in drawing up surveys, and implementing a degree of unification in the assessment criteria used in universities.

Keywords: teacher appraisal, teaching profession, undergraduate student, further education of teachers, teacher behaviour, university.

Resumen

Actualmente, las encuestas de satisfacción de los estudiantes universitarios sobre la docencia recibida están implantadas en la mayoría de las universidades de los países desarrollados. Existe un amplio debate sobre la pertinencia y verdadera finalidad de esta valoración.

Con el fin de coadyuvar al diseño de un instrumento de medición verdaderamente útil y que realmente incida en una mejora de la docencia universitaria, hemos analizado las encuestas de satisfacción que se realizan en las universidades públicas españolas.

A tal efecto, hemos desarrollado una investigación de enfoque mixto, basada en el análisis de contenido. El corpus de análisis han sido las universidades públicas españolas (N=47). Hemos codificado un total de 711 ítems en dimensiones y categorías temáticas, en función del aspecto que evalúa cada ítem, siguiendo un proceso inductivo basado en un criterio semántico. La fiabilidad entre codificadores fue calculada mediante alpha de Krippendorff, usando el programa estadístico SPSS y el macro de Hayes para variables nominales. Para el análisis de contenido realizamos una codificación humana utilizando el programa de análisis cualitativo atlas.ti.

Entre los resultados más significativos constatamos la existencia de: aspectos que no dependen directa ni indirectamente de la labor del docente y que son evaluados haciendo media con el resto de variables; otros, que no son responsabilidad directa del docente; algunos aspectos relevantes para la docencia que quedan sin evaluar; y, en la mayoría de los casos, una formulación mayoritaria de ítems cerrados que limitan la obtención de una valiosa información a utilizar para mejorar la actividad del profesorado. Concluimos afirmando la necesidad de rediseñar estos instrumentos, teniendo en cuenta tanto la perspectiva del alumnado como la de los docentes en la elaboración de las encuestas, y propugnado cierta unificación de los criterios de evaluación empleados en las universidades.

Palabras clave: evaluación del profesor, profesión docente, estudiante universitario de primer ciclo, perfeccionamiento de profesores, conducta del profesor, universidad.

Defining the issue

The Bologna Declaration (1999), the cornerstone and point of reference of the European Higher Education Area (EHEA), has transformed university education in Spain to favour European convergence. Research on higher education, especially teacher assessments (Fraile, López-Pastor, Castejón y Romero, 2013) has greatly increased following the obvious surge in evaluations monitoring the quality of university education (Arribas y Martínez, 2015).

The work of university teachers is of major importance in ensuring the suitable training of students for success in their education (TALIS, 2013), an increasingly complex task given the need to adapt to the continuous demands of modern society (Haigh, 2010). Furthermore, institutional policies geared mainly towards results have led to a notable increase in what is being asked of teachers (Leisyte, Enders y De Boer, 2009).

In Spain, the Organic Law of Universities brought about major changes in the accreditation and obligatory assessment of university teachers. At present, Spanish universities are obliged to evaluate the teaching competence of their staff. In 2007, the Spanish National Agency for Quality Assessment and Accreditation (ANECA) - in charge of improving the quality of the Spanish higher education system - implemented the DOCENTIA program to help universities design assessment mechanisms (ANECA, 2015). Universities still have individual autonomy to assess the teaching activity of their staff. Aspects to be considered include teaching publications, years of experience, innovation projects and the results of student surveys (González-Galán y Rodríguez-Patrón, 2014; ANECA, 2015).

Antecedents and theory

Teacher assessment questionnaires were first formally distributed among students at the University of Washington in 1920 (Guthrie, 1954) spreading in the English-speaking context in the 1960s before becoming popular worldwide (Spooren, Mortelmans y Thijssen, 2012). They are currently in use in many universities in developed countries, including Australia, the Netherlands and the United Kingdom, all of which have national surveys with mass participation (Sabri, 2013).

Received: 08-05-2018 Accepted: 22-03-2019

Although they were originally designed for educational purposes, they were first applied for decision-making concerning teaching staff in the 1970s (Galbraith, Merrill y Kline, 2012). Their results currently influence aspects such as teachers' wages, academic merits or teaching awards in the United States, Australia, Asia, and Europe (Hénard 2010), and have come to play a major part in the public administration of higher education (Sporn, 2011), becoming a highly standardised procedure (Jahangiri y Mucciolo, 2008). However, it has not been without controversy (Benton y Cashin, 2014). There are many drawbacks in addition to its advantages (Schneider, 2013). The frequent accusations of limited reliability and validity (Hornstein, 2017; Lee Hansen, 2014; Mittal, Gera, Batra, y Dharminder, 2015; Spooren, Brockx y Mortelmans, 2013), as well as the subsequent use of results (Casero, 2008), had great repercussions in the USA, where students are strongly influenced by these results when selecting courses and teachers. They also play a major part in the hiring, firing, promotion and salaries of teachers (Rosen, 2018).

To date, research on these surveys has provided no clear response as to their validity (Spooren et al. 2013). One of the main problems is identifying the content for evaluation. This is further complicated by the need to establish two cornerstones for the use of the content of these assessment surveys:

a. Establishing the aspects determining what a good teacher is. Detractors of this assessment system point out the lack of consensus on what the characteristics of a high-quality education ought to be (Nasser-Abu, 2017), as this depends on multiple factors (Moreno-Murcia, Silveira y Belando, 2015).

Several studies have analysed the content of different assessment systems using programs such as VERIFICA, AUDIT, ACREDITA and DOCENTIA. Factors highlighted include relationships and interactions with students, didactic methodology, knowledge of the subject matter, and assessment competences (González y López, 2010). However, the assessment indicators need to be adapted to different cultural contexts (Mittal et al. 2015) and specific educational problems.

b. Outlining the premise of satisfaction with the education received (Chacón, Pérez-Gil, Holgado y Lara, 2016). We must consider that students' personal expectations regarding learning negatively affect their satisfaction if not met. Pérez, Lozano, Gómez y Aguilera (2010) laid down the aspects defining satisfaction with the work of teachers as ability, command and accessibility of the teacher.

Received: 08-05-2018 Accepted: 22-03-2019

Assessments from Spanish public universities lay emphasis on student perceptions (González et al. 2010), but this aspect should go beyond mere description (Mai, 2017) in order to truly improve university teaching by identifying its weaknesses and strengths (Ordóñez-Sierra y Rodríguez-Gallego, 2015). For this reason, we have carried out a critical analysis of satisfaction surveys completed by students in Spain.

Design and methodology

Most of the research on student satisfaction surveys has focused on statistical reliability and validity (Spooren et al. 2012). We developed a mixed method study based on content analysis (Hsieh y Shannon, 2005) for the in-depth analysis of a series of themes and concepts (Cheng, Edwards, Darcy y Redfern, 2016), applying the phases proposed by Espín (2002):

- Pre-analysis. We selected student satisfaction surveys on the teaching work carried out in Spanish public universities. Of the 50 public universities we ruled out the two specific universities which only offer postgraduate studies[2], and a distance learning university[3], as they are not considered homogeneous with the rest of institutions studied. The body of analysis is therefore N=47 Spanish public universities.
- Selection of universities for analysis. We studied the items designed for the students, ruling out those aimed exclusively at postgraduate or laboratory students. Equally, we analysed those referring to teaching work, but not those referring to general data on the student (age, gender…) or group context which did not refer to teachers' work.

Following an inductive process, semantic criteria using themed measurements and categories were used to code the data extracted from the sample (N=711 items) depending on the aspect assessed in individual items. A total of 12 dimensions was subdivided into 32 themed categories.

(2) Universidad Internacional Menéndez Pelayo and Universidad Internacional de Andalucía.
(3) UNED.

Received: 08-05-2018 Accepted: 22-03-2019

In order to ensure maximum reliability for coding, we consulted seven independent experts to evaluate the relevance and clarity of the variables using a subsample of 10 surveys. The reliability of coders was calculated using Krippendorff's alpha, SPSS statistical software, and the Hayes macro (Hayes and Krippendorff 2007) for nominal variables.

TABLE I. Calculating Krippendorff's alpha

	Alpha	LL95%CI	UL95%CI	Units	Observers	Pairs
Nominal	.8684	.7306	.9701	12	7	273

Source: Authors' own

The results in table 1 show very high reliability between coders (α = 0.86)[4]. These measurements are used to define and analyse categories, as recorded in table 2.

TABLE II. Analysis dimensions and categories

Dimensions	Definition	Categories
Information provided to students by the teacher	Communication on aspects relating to curriculum which allows students to learn about the education project they are to take part in.	
Teacher planning	Project of the teaching activity that is to be carried out.	Fit to the established planning Teaching organisation
Contact hours	Guidance action from teachers to students in educational, personal and professional aspects.	Attention received Compliance with timetable Usefulness Accessibility/availability
List of contents	Connection between knowledge, abilities, skills and attitudes contributing to the attainment of the objectives of the subject matter.	Theory and practice Coordination of content in a specific subject and others Coordination between teachers of the same subject

(4) A result > 0.8 means major reliability between coders (Krippendorff, 2018).

Received: 08-05-2018 Accepted: 22-03-2019

Teacher explanations	Explanations from the teacher which facilitate perception of the subject matter content.	Clarity of explanations Use of examples Interest of the teacher in their explanations being understood Command of the subject matter Resolution of doubts
Methodology used	Procedure established to guide student learning on the subject matter.	Usefulness of didactic resources Usefulness of activities Methodological adaptation to the characteristics of the subject matter. Availability of resources
Motivation	Encouragement from the teacher to prompt student interest in the subject.	
Class atmosphere	Set of conditions or circumstances during the class which make the subject matter more approachable.	Climate of participation Treatment of students Encouragement of reflection and critical thinking
Dedication to the subject on the part of students	Time and effort invested by students in order to master the subject.	
Assessment	Calculation of the learning and education of the students as a result of the educational process.	Information offered Adaptation of the procedure to the subject matter Fit to the specifications of the teachers' guide Usefulness to learning
Global satisfaction	Enjoyment or complacency of students regarding the subject matter and the task carried out by the teacher.	With the teaching work With the learning acquired With the subject With satisfaction of teaching obligations
Open questions	Matters which allow the surveyed student to answer freely.	Observations/comments Positive aspects Negative aspects or aspects to be improved

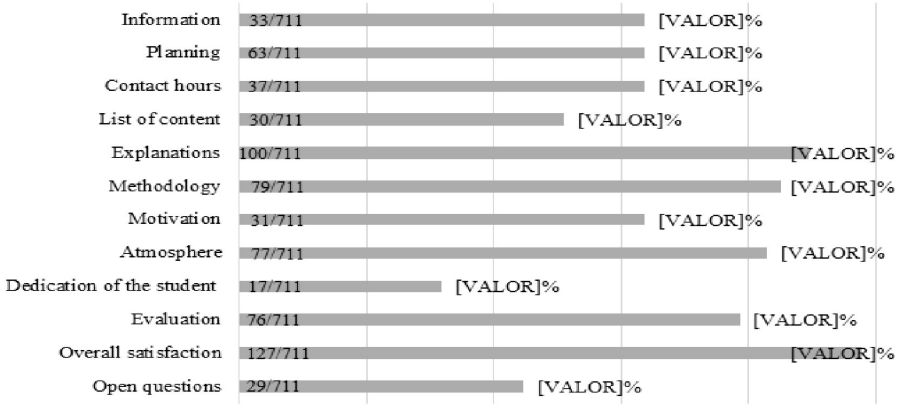
Source: Authors' own

- Use of the material. For purposes of content analysis, human coding was carried out using atlas.ti qualitative analysis software for the categorisation process. This allowed the survey items to be reduced to analysis units (Armborst, 2017).
- Finally, the data and results obtained were systematised and interpreted based on the proposed objectives.

Results

In the analysis and description of the results we have considered the presence of the dimensions and their respective categories presented in percentages in satisfaction surveys (N=47) as well as the number of items for individual evaluation (N=711), the frequency of which (n/711) is included in graph 1. Overall student satisfaction, teacher explanations and the methodology used by teachers are the most common aspects of the surveys and have the highest number of items. In contrast, the list of contents, open questions and dedication of students to the subject matter appear with the least frequency.

GRAPH I. Aspects assessed in student satisfaction surveys



Source: Authors' own

Overall satisfaction

In this aspect, the most assessed category is 'general satisfaction with teaching work' (91.5% of surveys; 70/711 items). In this category it is worth highlighting the item: *How do you think your classmates value the work of the teacher in this subject?* (University of Murcia). This does not reflect the direct opinion of the student surveyed but rather

Received: 08-05-2018 Accepted: 22-03-2019

presupposes that students know this opinion. It does not take into account the circumstances and relationships between classmates, so that the responses to this item can lead to biased results.

'Compliance with teaching obligations' is assessed in 25.5% of universities (27/711). Most of the items question whether the teacher attends class and respects the established timetable. Some universities use different items to evaluate individual variables: *Teaches class in the established timetable*, and *Regularly attends class* (U. Pablo Olavide, Malaga, Jaén, Huelva and Cadiz). However, many surveys include different variables to be assessed within a single item, which should be considered separately for correct evaluation. This erroneous formulation complicates the response of students, who have to give single answer for a series of questions. Some examples of this would be: *Full advantage is taken of class time: timetables are respected and students' self-education is facilitated* (University of Navarra); *Attends class, but in the event of non-attendance class is justified, replaced or recovered* (University of Zaragoza); and, *Has suitably satisfied obligations as a teacher (teaching plan, subject curriculum, attendance, delivery of marks, etc.* (University of Barcelona).

The least assessed categories in this aspect are 'satisfaction of students with the subject' (25.5%, 16/711), and 'with acquired learning' (25.5%, 14/711). Many of the aspects which make up the subject are set by the Ministry of Education, Culture and Sports, while in many cases the objectives, competences, content, bibliography and evaluation systems are dictated by the coordinator of the subject and not by the teachers.

Teacher explanations

'Clarity of explanations' is the second most assessed category from all universities (72.3%; 37/711). Once again, different independent variables are assessed simultaneously in a single item: *Organises classes well, offering clear explanations and making it easier to understand the subject*[5] (University of Santiago de Compostela), even if clear explanations and well-organised classes do not always go hand-in-hand.

(5) As two categories are contemplated 'organisation of teaching' from the 'teacher planning' dimension is also included.

Received: 08-05-2018 Accepted: 22-03-2019

'Resolution of doubts' is assessed in 63.8% of the universities (30/711). Within this category it is worth highlighting the item: *Shows an interest in the extent to which explanations are understood while offering any clarifications proposed*[6] (University of Almería).

'Interest of teachers in their explanations being understood' is assessed in 29.8% of universities (13/711). We believe this aspect to be highly appropriate in the assessment of any teaching activity, as it makes it possible to establish whether the teaching-learning process is being conducted correctly, leading to optimal educational results.

The least assessed categories are 'teacher's command of the subject matter' (27.6% (13/711), and the 'presentation of examples' in explanations (17%, 8/711).

Methodology

The methodology has been subdivided into four categories, the most widely assessed of which is 'usefulness of didactic resources' (78.7%; 55/711). Two specific wordings are worth highlighting: *the teacher provides additional documentation (plans, graphs and diagrams, etc.) to the content* (University of Las Palmas) and *The teacher encourages the use of additional resources (bibliographical or otherwise) to those used in the classrooms* (University of La Coruña). The fact that supporting documentation or additional resources are used does not necessarily imply their usefulness in student learning, nor do they represent better teaching.

The 'usefulness of the activities' carried out is evaluated in 46.8% of the universities surveyed (23/711). This category covers clearly improvable categories including *The activities set by the teacher in non-presential hours are useful to my learning* (University of Cartagena), assessing a type of activity which not all teachers use or are obliged to use, and that do not imply better teaching. Furthermore, the fact that this item does not allow for the response 'I don't know' ('DK') suggests that if the teacher does not carry out this kind of work students tend to rate this variable negatively, even if the activity is not carried out.

(6) As two categories are contemplated 'interest in the level of understanding of their explanations' is also included.

Received: 08-05-2018 Accepted: 22-03-2019

Among the least evaluated categories we find 'availability of resources' (8.5%, 7/711) where the following items are of note: *As a student are you able to easily access the facilities, services, material resources and technological equipment and bibliographical archives of the University in accordance with the Degree Programme* (University of Las Palmas); and *The technical resources (aula net, photocopiers, etc.) available are sufficient to develop the work in this way outside the classrooms* (University of La Coruña). These items evaluate university services, not the work of the teachers.

The 'adaptation of the methodology to the characteristics of the subject matter' is only assessed in 8.5% of universities and in 6 items, including *The modes of teaching-learning fit: A. the features of the student group; B. the nature of the subject and C. our learning needs* (University of País Vasco). This establishes a single response option for each of the three cases possible, worded in a concrete and specific way and favouring a non-ambiguous assessment and interpretation.

Class atmosphere

In this case, the most assessed category is whether the teacher actively 'encourages a suitable atmosphere for work and participation' in the subject (74.5%; 45/711). However, some universities use three different items to evaluate class atmosphere and individual participation. In the University of País Vasco, the categories used are: the teacher encourages student participation in class; assessment of classroom atmosphere; and assessment of teacher communication with students.

Another category is 'teacher's treatment of the students' (48.9%; 27/711). Again, this is an item simultaneously assessing teachers' accessibility and how they respond to queries all in the same point. This hinders the response from students and its subsequent interpretation: *Teachers treat the students correctly, are amenable to suggestions and queries and respond to these adequately* (University of Navarra).

The least examined category in this aspect is the 'encouragement of reflection and critical thinking' among students on the part of the teacher (10.6%; 5/711).

Received: 08-05-2018 Accepted: 22-03-2019

Assessment

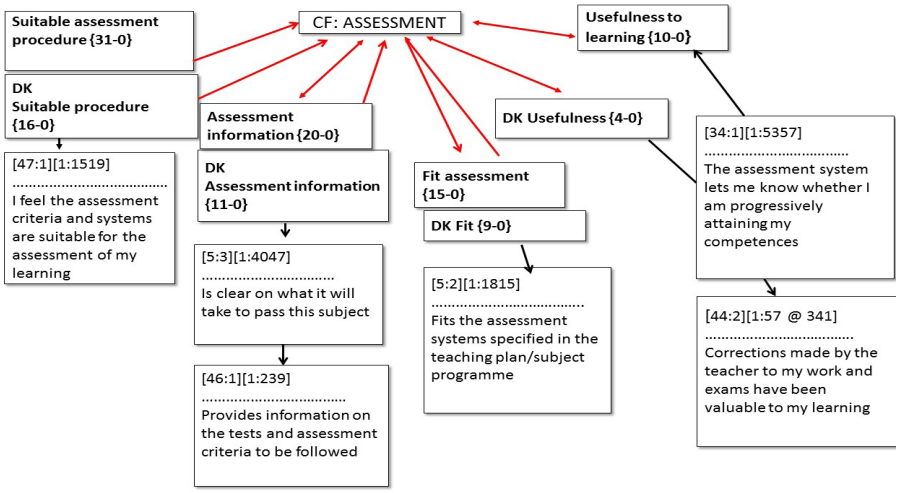
Over half of the surveys (53.2%; 31/711) assess whether the 'evaluation procedure is suited to the subject matter'. Among the numerous striking aspects of this category is the fact that just one university (Vigo) clarifies that this is only applicable in cases where some form of assessment or continuous assessment has been carried out. In order for students to express an opinion on this and other aspects relating to this assessment, they must first have been evaluated in some way.

Furthermore, most surveys simultaneously assess criteria and the evaluation system, variables which ought to be examined individually.

Thirdly, it is worth noting the item *Takes into consideration students' opinions when establishing the procedures to assess learning* (University of País Vasco). The assessment procedure is established before beginning work on the subject and the 'information received by students at the start of the academic years on the method for the assessment of the subject' is present in 38.3% of surveys (20/711).

The remaining categories which are less frequently assessed in this respect are the 'assessment system is suited to the specifications of the teacher's guide' (19.8%; 15/711); and the 'usefulness of the evaluation process for student learning' (21.3%; 10/711).

GRAPH II. Conceptual map of the 'Assessment' aspect



Source: Author's own

Received: 08-05-2018 Accepted: 22-03-2019

Only 3 of the 47 universities studied carried out the surveys online, after the exams or relevant final assessment. This timing is critical. Even if the teachers have informed the students about the marking system and criteria, if the satisfaction survey is carried out prior to assessment it cannot really be established whether these criteria and system are suited or useful to learning or have been respected. This is further complicated - as graph 2 shows - by the fact that only 40 of the total 76 items for this aspect include the answer option 'DK', affecting the veracity of the information obtained.

Teacher planning

All the surveys assessing this point evaluate the 'fit of the subject matter to the planning as established by the teacher' (37/711). Here it is worth noting the item *The programme taught up till now suggests that it will be developed in full* (University of La Rioja), which evaluates a supposition, not a fact. In addition, the item *The programme of the subject matter is taught in full* (University of Navarra) cannot be suitably assessed if the survey takes place before the subject is completed.

The category 'organisation of teaching', assessed in 48.8% of universities (26/711), notably includes 4 different items for the evaluation of the preparation of teaching activity (University of Vigo): *I believe the teacher thoroughly prepares and organises the activities and tasks in (answer only when applicable): a. theory classes; b. practical classes; c. others (exercise solving, case studies, seminar…); d. individual student work (supervised work, projects, documentation…).*

Contact hours

'Attention received from the teacher' (21.3%; 11/711) and 'compliance with the timetable' (21.3%; 10/711) are the most assessed categories in this respect. In the case of the latter the item: *The timetable for contact hours is adequate* (University of Seville) ought to be noted. Many teachers have their group classes in morning and afternoon timeslots. If also considering the personal and/or professional commitments of individual teachers, a timetable that will suit the needs of all students is unlikely to

Received: 08-05-2018 Accepted: 22-03-2019

be established. The fact that a teacher sets contact hours does not imply that they are not accessible or cannot adapt to a timetable other than that established. The wording of the item is not suitable for the evaluation of the availability and accessibility of contact hours with the teacher.

The categories assessed with the least number of items in this respect are the 'usefulness of contact hours', and the 'accessibility/availability of teachers in contact hours' (in both cases 17%, 8/711). The nuances of some of the surveys stating that these items should be completed only if the contact hours have been used are praiseworthy. An example of this is *Degree of compliance with the contact hours set by the teacher (if never attended, please leave blank)* (University of La Rioja). We find this to be an appropriate clarification as any students who have not made use of contact hours are in no position to formulate an objective assessment. Therefore, there is a risk of prompting a random and unsubstantiated response from the student surveyed. It is striking that of the total of items assessing contact hours (37), 12 do not include this specification or at least the option 'DK'.

Information offered to the students

Rather surprisingly, this aspect is one of the least assessed in the overall number of items (33/711), despite its importance in the proper development of the subject and the correct assessment of the teacher.

Most surveys resort to the item *The teacher provides information on the different aspects of the teaching programme (objectives, activities, curriculum content, methodology, bibliography, assessment systems, etc.),* grouping different aspects under a single wording as can be observed, hindering both the assessment carried out by the students and the feedback provided to the teachers. In order to obtain an accurate evaluation, it would be necessary to establish different items for each aspect to be assessed, for example, *The teacher provides clear information on the objectives of the subject* (U. Complutense Madrid).

Motivation

In this aspect it is worth noting an item which could be highly suitable and is surprisingly only used in one university: *The teacher makes me see the*

Received: 08-05-2018 Accepted: 22-03-2019

importance of this subject and its relevance to my education (University of Santiago de Compostela). This is considered a very appropriate assessment as one of the aims of the subject must be for students to find meaning in the education planned by the teacher, making it significant for education and subsequent professional careers.

List of contents

The most repeated category in this dimension is 'coordination between the theory and practice of the subject' (29.85; 14/711). This is followed by 'coordination/relationship between content' (21.3%; 11/711), which values the link between the themes of the same subject matter, and the relationship or overlap with content from other subjects. In the case of the latter, responsibility is shared between teachers assessed and some of their colleagues. It is worth highlighting the item *The education received previously helps to attain the objectives of this subject* (University of Las Palmas), assessing an aspect that is of great importance, but in no way the responsibility of the teacher of the subject. Firstly, the order and distribution of subjects by year level is established by the university itself with no input from teachers in terms of contributions and/or responsibilities. Furthermore, education previously received by students is the competence of previous teachers, not those being assessed.

'Coordination between teachers of the subject' is included in 10.6% of surveys (5/711). The item *The work carried out by this teacher fits satisfactorily with the work carried out by other teachers* (University of Valencia) should be noted, as it values not only the teaching activity of a specific teacher, but that of all the teachers involved.

Open questions

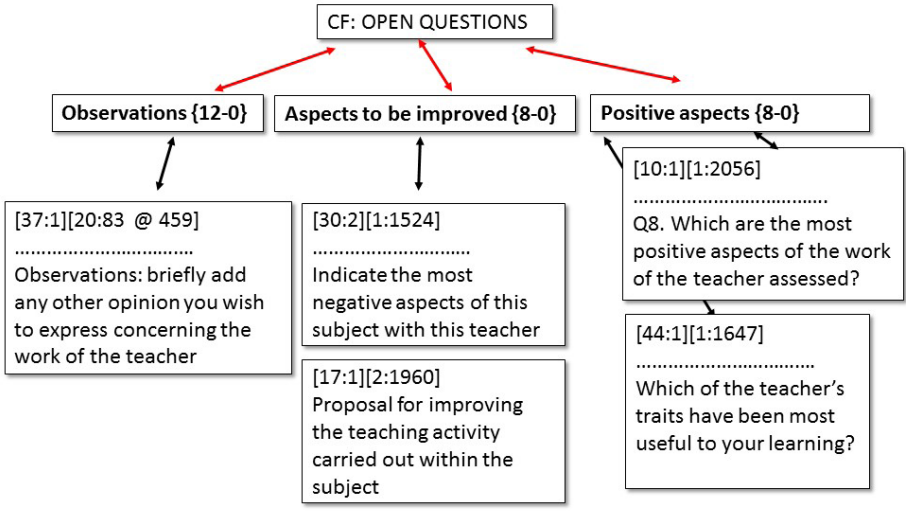
Open questions are one of the two least assessed aspects. As graph 3 shows, the least frequent category is 'observations or comments' (23.4%%; 12/711).

Some surveys include open questions where the students can add any comment, allowing them to assess aspects not reflected in the standard items. In some it is possible to identify 'negative aspects or proposals' in

Received: 08-05-2018 Accepted: 22-03-2019

order to improve teaching in the subject (17%; 9/711) (*State, if applicable, the aspects which the teacher of this subject ought to improve (offensive answers will not be taken into consideration)*, University of León); while in others it is possible to highlight the 'positive aspects' of the teacher (17%; 8/711).

GRAPH III. Conceptual diagram of 'Open questions'



Source: Author's own

Dedication of students to the subject

This point assesses whether the level of work completed by the students is in keeping with the subject and the number of credits established. Among these, the item *How difficult do you find this subject compared to others you study or have studied?* (University of La Rioja) should be noted, as the assessment depends on the comparison unit chosen by the students to make the analogy, and not on a standard criterion. Furthermore, we consider that each subject is intrinsically different as is the degree of difficulty, without taking into account the multiple variables

Received: 08-05-2018 Accepted: 22-03-2019

influencing students' attitude to the subject. These aspects are not the direct consequence of the work of the teacher.

Discussion of results and conclusions

Most of the surveys include items assessing aspects on which students have no information or sufficient experience to provide a reliable assessment. These must only be answered when students have suitable knowledge of the aspect to be assessed, as in the University of León, or introducing the answer option 'DK' in all items[7]. Judgement based on conjecture and not known facts is to be avoided. For example, in most surveys students are expected to provide opinions on the suitability of criteria and evaluation systems for teachers before completing the subject. They have to indicate whether these comply with the student prospectus, adapt to the subject matter and whether they have been of use for learning and assessment.

We recommend answering the surveys after the assessment process is completed but before the publication of the final marks for the subject, so that these cannot subjectively influence the assessment of the teacher/examiner. Deliberate revenge for low marks or a subtler process such as cognitive dissonance could lead to low ratings (Boysen, 2008). It is possible to carry out the surveys after the final assessment of the subject in only 13 (27.6%) universities, although they are usually handed out earlier in order to prevent a predictable drop in participation. In order to guarantee higher participation, students could be encouraged with a participation certificate or something similar. In some universities, such as Simon Fraser in Canada, draws are held for the students who have completed assessment questionnaires[8] (Simon Fraser University, 2017). Although the bias of the students not completing the survey is unknown (Wolbring, 2012) they tend to be low-performing and dissatisfied students (Dobkin, Gill y Marion 2007).

Various studies conclude that the online surveys have significantly lower ratings than those carried out in class (Nowell, Gale, y Handley, 2010). Therefore, if students are able to complete these online they should

(7) Only 51% of surveys and 54.4% of items offer this option.
(8) In the 2016-2017 academic year the prizes were an iPad, and gift vouchers to the value of 100$ and 50$.

Received: 08-05-2018 Accepted: 22-03-2019

be given the chance to complete them in class (Treischl y Wolbring, 2017).

We have observed that the assessments are aimed exclusively at compulsory, basic or elective subjects. However, this does not apply to work tutored by the teacher in external placements or the end-of-degree project, which require major organisation, preparation, tutoring, attention, guidance and review. This omission results in these teaching activities being undervalued, as good or bad practice of the teacher in this respect would have no repercussion and teachers would be deprived of information that would be valuable for improving their work.

As stated by González et al. 2010 some of the items are irrelevant or poorly thought out. They are often unsuitable or the wording is poorly formulated, including more than one variable for evaluation. This deprives the teachers of accurate information which would allow them to learn more about their teaching and improve it. In fact, drawing up a tool made up of well-formulated and presented items is crucial to ascertaining the level of satisfaction among students regarding the teachers' work, with results impacting their professional careers (Moreno-Murcia et al. 2015).

In our opinion, both teachers and students should be surveyed on the aspects to be evaluated in these questionnaires. The relevance and suitability of items should also be verified before being included in the final version, especially bearing in mind the lack of consensus among teachers and students on what constitutes high-quality education (Nasser-Abu, 2017) and an efficient teaching-learning process (Bosshardt y Watts, 2001). There is only one instance, in the University of La Rioja, where students are asked about their knowledge of the teachers' assessment system and whether they consider the survey items suitable.

The most widespread categories in the surveys, and those considered most important, are the usefulness of didactic resources, the atmosphere of participation promoted by the teacher and the clarity of explanations. Therefore, other aspects such as the encouragement of reflection and critical thinking, usefulness and accessibility of contact hours, open questions to compile opinions on positive aspects, and suggestions for improvement of teaching activity are in the minority and considered less important.

The differences between some subjects and others which are more prone to the use of didactic media and support considered important a priori could lead to disagreement among teachers. This would partly be

the result of different interpretations of freedom of teaching, as laid down by the Spanish Constitution (art. 20.1.c.) and the Charter of Fundamental Rights of the European Union (art. 13).

It is striking that from a total of 711, only 29 (4%) open items are counted. We agree with Van Gelder's (2017) assessment that standard questions are very general, require a very simple rating from students and do not reveal a precise opinion. The inclusion of open questions is advised so that students may express their opinions freely, thus enabling more accurate and useful data to be collected.

We propose the creation of a homogeneous teacher evaluation system adapted to the European Higher Education Area (Sabri, 2013), respecting the autonomy of individual universities and including a category to examine the particular characteristics of each university (Mittal et al. 2015). In 2007 ANECA implemented the DOCENTIA program to support the assessment of university teachers' activity, although it is not yet used in all Spanish universities (ANECA, 2015). This initiative contemplates each university controlling the timeframe of the student satisfaction surveys and the evidence compiled from them. This would cover items relating to the self-assessment of the teacher, curricular teaching merits, reports from academic supervisors and result ratings from the subjects taught.

Although the assessments provided by the students are a means to improving the education they receive (Rantanen, 2013), we do not agree with teachers being categorised and being made responsible for their respective institutions (Baez, 2014). Recent meta-analysis of multisector studies did not report significant correlations between student assessments in these surveys and their learning (Spooren y Christiaens, 2017).

These surveys may reflect a biased view of reality (La Serna, Becerra, Beltrán y Hongrui, 2014), as the student is not necessarily aware of the contextual variables affecting teaching, such as the resources available to teachers in class (Díaz, 2015). There is also an erroneous starting premise that the anonymity of students, regardless of whether they attend class or not, guarantees a sincere response and an objective and accurate assessment of the teacher's individual teaching skills. This point is highly debatable, given the wealth of aspects which can significantly affect student satisfaction with teachers, including the perception and expectations of individual students regarding teaching-learning, connection with the teacher (Feistauer y Richter, 2017), and the charisma or personality of the teacher providing a clear 'halo' effect (Mittal et al. 2015).

Received: 08-05-2018 Accepted: 22-03-2019

Furthermore, there is no consensus on which competences ensure that teachers provide a high-quality education. Although there is a belief that good teachers require their students to make an effort, students do not always agree, and this may result in negative assessments (Braga, Paccagnella y Pellizzari, 2014). In some universities in the Netherlands this has led teachers to question the advisability of reprimanding or being strict with their students, given the risk of being given lower scores in their assessment (Van Gelder, 2017). Some studies also state that the standard satisfaction questionnaires used are not suitable for evaluation of subjects or teachers whose objectives and methodology differ significantly from the conventional approach (Lee Hansen, 2014).

Contributions

Teacher satisfaction surveys are an assessment tool to be used, always bearing in mind their limitations and combining them with other sources of information, such as systematic observation and teacher self-assessment, thus allowing the information obtained to be triangulated (Montoya y Ramírez, 2013; Salazar, 2008). Teacher assessment exclusively from students is a unidirectional process. However, given the bidirectional nature of the teaching-learning process, many authors consider that the opinions of both students and teachers should be kept in mind (Álvarez, García, Gil, Romero y Correa, 2002; Gutiérrez-García, Pérez-Pueyo, Pérez-Gutiérrez y Palacios-Picos, 2014). In keeping with Casero (2011) we agree that teacher satisfaction surveys should be drawn up with the consensus of both parties concerned.

The popularity and widespread use of these surveys (Spooren et al. 2013) is mostly due to the simple measurement process they offer. It is extremely easy to compare the numerical results between teachers, departments and faculties. However, the lack of consensus among researchers regarding their validity as a gauge (Hornstein, 2017) and their limitations lead us to conclude that they should not be used as the basis for decision-making concerning careers, as is the case in the USA and Canada (Boring, 2016). They should however be used to provide feedback to teachers and institutions in order to improve teaching and in turn the university education provided.

Received: 08-05-2018 Accepted: 22-03-2019

Limitations of the study

The main limitation of this study is the restrictions characteristic of the chosen method. The coding process intrinsic to content analysis confers subjectivity to the study. However, if the prior analysis for reliability of the tool is fully complied with, the coding phase becomes a mere systematic administration of the decisions taken. In this case therefore, high reliability between external coders has been observed in addition to the comprehensive categorisation of aspects, definition and analysis categories. Nevertheless, in order to ensure greater agreement and increase reliability in the coding process, the categorisations were carried out by two coders on two separate occasions each.

Bibliography

Álvarez, V., García, E., Gil, J., Romero, S. y Correa, J. (2002). Enseñanza en la universidad. *Revista de Educación,* nº 328, 303-323.

ANECA (2015). *Programa DOCENTIA*. Madrid: ANECA. Retrieved from http://www.aneca.es/Programas-de-evaluacion/DOCENTIA/Documentacion-del-programa

Armborst, A., (2017). Thematic Proximity in Content Analysis. *SAGE Open*, vol. 7, nº 2, 1–11. doi: 10.1177/2158244017707797

Arribas, J. y Martínez, C. (2015). El programa AUDIT de la ANECA y las normas internacionales ISO 9000. *Educación XX1*, vol. 18, nº 2, 375-395. doi: 10.5944/educxx1.14609

Baez, B. (2014). *Technologies of government: politics and power in the information age*. Charlotte: Information Age Publishing.

Benton, S. y Cashin, W. (2014). Student ratings of instruction in college and university courses. In M. Paulsen (Ed.), *Higher education: Handbook of theory and research* (279-326). Netherlands: Springer.

Boring, A., Ottoboni, K. y Stark, P. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *Science Open Research*, vol. 0, 1-11. doi: 10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1

Bosshardt, W. y Watts, M. (2001). Comparing Student and Instructor Evaluations of Teaching. *Research in Economic Education*, nº 32, 3 -17.

Boysen, G. (2008): Revenge and Student Evaluations of Teaching. *Teaching of Psychology*, vol. 35, nº 3, 218-222. doi: 10.1080/00986280802181533

Braga, M., Paccagnella, M. y Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, nº 41, 71–88. doi: 10.1016/j.econedurev.2014.04.002

Casero, A. (2008). Propuesta de un cuestionario de evaluación de la calidad docente universitaria consensuado entre alumnos y profesores. *Revista de Investigación educativa*, vol. 26, nº 1, 25-44.

Chacón, S., Pérez-Gil, J., Holgado, F. y Lara, A. (2011). Evaluación de la calidad universitaria: validez de contenido. *Psicothema*, vol. 13, nº 2, 294-301.

Cheng, M., Edwards, D., Darcy, S. y Redfern, K. (2016). A tri-method approach to a review of adventure tourism literature. *Journal of Hospitality & Tourism Research*, nº 3, 1-24. doi: 10.1177/1096348016640588

Díaz, W. (2015). Formación del profesorado universitario, evaluación de la actividad docente y promoción profesional (Tesis doctoral). UNED. Retrieved from: http://e-spacio.uned.es/fez/eserv/tesisuned:Educacion-Wbdiaz/DIAZ_MARTIN_William_Tesis.pdf

Dobkin, C., Gil, R. y Marion, J. (2007). Causes and consequences ok skipping class in college. *Jour or economic education*, vol. 3, nº 3, 251-256.

Espín, J. (2002). El análisis de contenido: una técnica para explorar y sistematizar la información. XXI. *Revista de Educación,* nº 4, 94-105.

Feistauer, D. y Richter, T. (2017). How reliable are students' evaluations of teaching quality? *Assessment & Evaluation in Higher Education*, vol. 42, nº 8, 1263-1279. doi: 10.1080/02602938.2016.1261083

Fraile A., López-Pastor, V., Castejón, J. y Romero, R. (2013). La evaluación formativa en docencia universitaria y el rendimiento académico del alumnado. *Aula Abierta*, vol. 41, nº 2, 23-34.

Galbraith, C., Merrill, G. y Kline, D. (2012). Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business related classes? *Research in Higher Education*, nº 53, 353-374. doi: 10.1007 / s11162-011-9229-0

González, I. y López, A. (2010). Sentando las bases para la construcción de un modelo de evaluación a las competencias docentes del profesorado universitario. *Revista de Investigación Educativa*, vol. 28, nº 2, 403-423.

Received: 08-05-2018 Accepted: 22-03-2019

González-Galán, M. y Rodríguez-Patrón, P. (2014). La evaluación del profesorado universitario en España. *Revista de educación*, nº 366, 136-164. doi: 10.4438/1988-592X-RE-2014-366-279.

Guthrie, E. (1954). *The evaluation of teaching: a progress report*. Washington: University Washington.

Gutiérrez-García, C., Pérez-Pueyo, Á., Pérez-Gutiérrez, M. y Palacios-Picos, A. (2014). Percepciones de profesores y alumnos sobre la enseñanza, evaluación y desarrollo de competencias en estudios universitarios de formación de profesorado. *Cultura y Educación*, vol. 23, nº 4, 499-514. doi: 10.1174/113564011798392451

Haigh, A. (2010). *Enseñar bien es un arte. Sugerencias para principiantes*. Madrid: Narcea.Hayes, A, y Krippendorff, K. (2007). Answering the call for standard reliability measure for coding data. *Communication methods and measure*, vol. 1, nº 1, 77-89. doi: 10.1080/19312450709336664

Hénard, F. (2010). *Learning our lesson*. Paris: OECD.

Hornstein, H. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance, *Cogent Education*, vol. 4, nº 1. doi: 10.1080/2331186X.2017.1304016

Hsieh, H. y Shannon, S. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, vol. 15, nº 9, 277-1288. doi: 10.1177/1049732305276687

Jahangiri, L. y Mucciolo, T. (2008). Characteristics of effective classroom teachers as identified by studentsand professionals. *Journal of Dental* Education, vol, 72, nº 4, 484-493.

Krippendorff, K. (2018). *Content analysis*. Calofornia: SAGE Publications

La Serna, K., Becerra, A., Beltrán, A. y Zhang, H. (2014). La relación de las encuestas de evaluación docente con el rendimiento académico. *Revista Iberoamericana de Evaluación Educativa*, vol. 7, nº 2, 105-115.

Lee Hansen, W. (2014). Rethinking the student course evaluation. *Liberal Education*, vol. 100, nº 3. Retrieved from: https://www.aacu.org/publications-research/periodicals/rethinking-student-course-evaluation

Leisyte, L., Enders, J. y De Boer, H. (2009). The balance between teaching and research in Dutch and English universities in the context of university governance reforms. *Higher Education*, vol. 5, nº 5, 619-635. doi: 10.1007/s10734-009-9213-1

Received: 08-05-2018 Accepted: 22-03-2019

Mai, A. (2017). Pedagogical Innovation in Canadian higher education. *Studies in Educational Evaluation*, nº 54, 71-82. doi: 10.1016/j.stueduc.2016.11.001

Mittal, S., Gera, R., Batra, D. y Dharminder, K. (2015). Evaluating the validity of Student Evaluation of Teaching Effectiveness (SET) in India. *Education and Training*, vol. 57, nº 6, 623-638. doi: 10.1108/ET-06-2013-0072

Montoya, J. y Ramírez, M. (2013). Modelos de evaluación docente en la universidad. *Rhec, vol.* 16, nº 16, 273-300.

Moreno-Murcia, J., Silveira, Y. y Belando, N. (2015). Questionnaire evaluating teaching competencies in the university environment. *New Approaches in Educational Research*, vol. 4, nº 1, 54-61. doi: 10.7821/naer.2015.1.106.

Nasser-Abu, F. (2017). Teaching in higher education. *Studies in Educational Evaluation*, nº 54, 4-12. doi: 10.1016/j.stueduc.2016.10.006

Nowell, C., Gale, L. y Handley, B. (2010). Assessing faculty performance using student evaluations of teaching in an uncontrolled setting. *Assessment & Evaluation in Higher Education*, vol. 35, nº 4, 463-475. doi: 10.1080/02602930902862875

Ordóñez-Sierra, R. y Rodríguez-Gallego, M. (2015). Docencia en la universidad. *Bordón,* vol. 67, nº 3, 85-101. doi: 10.13042/Bordon.2015.67305.

Pérez, J., Lozano, J., Gómez, M., y Aguilera, A. (2010). Diseño de un instrumento para la evaluación de la satisfacción de la formación recibida de las diferentes asignaturas correspondientes al plan de estudios del grado de Psicología de la Universidad de Sevilla. *Revista de Enseñanza Universitaria*, nº 36, 45-61.

Rantanen, P. (2013). The number of feedbacks needed for reliable evaluation. *Assessment & Evaluation in Higher Education*, nº 38, 224–239. doi:10.1080/02602938.2011.62547

Rosen, A. (2018). Correlations, trends and potential biases among publicly accessible web-based student evaluations of teaching. *Assessment & Evaluation in Higher Education*, vol. 43, nº 1, 31-44. doi: 10.1080/02602938.2016.1276155

Sabri, D. (2013). Student evaluations of teaching as 'Fact-Totems'. *Sociological Research Online*, vol. 18, nº 4, 1-10. doi: 10.5153/sro.3136

Salazar, J. (2008). Diagnóstico preliminar sobre evaluación de la docencia universitaria. *Revista Iberoamericana de Evaluación Educativa*, vol. 1, nº 3, 68-84. doi: 10.15366/riee

Schneider, G. (2013). Student evaluations, grade inflation and pluralistic teaching. *Forum for Social Economics*, nº 42, 122–135. doi:10.1080/07360932.2013.771128

Simon Fraser University (2017). *Student evaluation of teaching and courses*. British Columbia, Simon Fraser University. Retrieved from: https://www.sfu.ca/setc.html

Spooren, P., Brockx, B., y Mortelmans, D. (2013). On the validity of student evaluation of teaching. *Review of Educational Research*, nº 83, 598-642. doi: 10.3102/0034654313496870

Spooren, P., Mortelmans, D., y Thijssen, P. (2012). 'Content' versus 'style'. *British Educational Research Journal*, nº 38, 3-21. doi:10.1080/01411926.2010.523453

Spooren, P. y Christiaens, W. (2017). I liked your course because I believe in (the power of) student evaluations of teaching (SET). *Studies in Educational Evaluation*, vol. 54, 43-49. doi: 10.1016/j.stueduc.2016.12.003

Sporn, B. (2011). Governance and administration. En J. Forest y P. Altbach (Eds.), *International handbook of higher education* (141-157). Netherlands: Springer.

TALIS (2013). *Estudio internacional de la enseñanza y el aprendizaje*. Madrid: Ministerio de Educación, Cultura y Deporte. Retrieved from: https://www.oecd.org/edu/school/Spain-talis-publicaciones-sep2014_es.pdf

Treischl, E. y Wolbring, T. (2017). The causal effect of survey mode on students' evaluations of teaching. *Research in Higher Education*, vol. 58, nº 8, 904-921. doi: 10.1007/s11162-017-9452-4

Van Gelder, L. (2017). *Teachers evaluation: you need an elephant skin*. Netherlands: Tilburg university. Retrieved from: https://universonline.nl/2017/06/19/teachers-evaluation-need-elephant-skin

Wolbring, T. (2012). Class attendance and students' evaluations of teaching. *Evaluation Review*, vol. 36, nº 1, 72-96. doi: 10.1177/0193841X12441355

Contact address: Vicente Llorent-Bedmar. Universidad de Sevilla. Facultad de Ciencias de la Educación, Departamento de Teoría e Historia de la Educación y Pedagogía Social. C/ Pirotecnia s/n, 41013, Sevilla. E-mail:llorent@us.es

Received: 08-05-2018 Accepted: 22-03-2019

Leadership in the new Spanish Army Education System[1]

Liderazgo en el nuevo sistema de educación del Ejército de Tierra de España

DOI: 10.4438/1988-592X-RE-2019-385-419

Natalia Utrero González
Francisco José Callado Muñoz
Montserrat Aiger Valles
Centro Universitario de la Defensa Zaragoza

Abstract

The Spanish army has recently implemented a new integrated educational curriculum to train officers by bringing together excellent and up-to-date technical knowledge, with a special emphasis on leadership, careful physical preparation and a solid moral foundation. The new curriculum is part of a carefully renovated human resources policy to develop a new type of leadership capable of adapting and performing successfully in the flexible and uncertain environments where national armies operate. The objective of the paper is to carry out an analysis of the implementation of the new educational system following the framework of Wright and Nishii (2006) that allows evaluating the implementation of new initiatives aimed at changing human resources policy. The main axes of the new system are critically analyzed and, based on this analysis; its implementation is evaluated by means of a survey carried out on its users. The survey is made to the first promotion that has been formed in the new curriculum once completed their training at the University Center of Defense and before going to the special academies. Information is required regarding three fundamental areas:

[1] Funded by the Spanish Ministry of Education and Science, ECO2016-76255-P y ECO2017-86305-C4-2-R; Reference research group COMPETE (S52_17R) and Grupo Psicología: procesos cognitivos y social (S24_17R) Gobierno de Aragón -España- and FEDER 2014-2020 "Construyendo Europa desde Aragón"; Universidad de Zaragoza, UZCUD2018-SOC-04 y UZCUD2018-SOC-O1; Centro Universitario de la Defensa, 2017-21 y 2018-12.

Received: 10-05-2018 Accepted: 29-03-2019

training at the center, internal practices and external practices. The results of the survey allow affirming that the design, communication and implementation of the curriculum is positively evaluated by the cadets. However, non-curricular training actions, such as external and internal practices and the development of leadership activities in real situations, which are considered complementary to formal education, receive a low evaluation, considering them not related to leadership. This evidence suggests aspects to be reviewed in the new education system with the aim of improving its effectiveness.

Key words: education, leadership, curriculum, transformation, human resources

Resumen

El ejército español ha implementado recientemente un nuevo currículum educativo integrado para formar a los oficiales aunando un conocimiento técnico excelente y actualizado, con un especial énfasis en el liderazgo, la preparación física cuidadosa y una sólida base moral. El nuevo plan de estudios es parte de una política de recursos humanos cuidadosamente renovada para desarrollar un nuevo tipo de liderazgo capaz de adaptarse y realizar un buen desempeño con éxito en los entornos flexibles e inciertos donde suelen operan los ejércitos nacionales. El objetivo del trabajo es realizar un análisis de la implementación del nuevo sistema educativo siguiendo el marco de Wright y Nishii (2006) que permite evaluar la puesta en marcha de nuevas iniciativas dirigidas a cambiar la política de recursos humanos. Se analizan de manera crítica los principales ejes del nuevo sistema y en base a este análisis se evalúa su implantación mediante una encuesta realizada a los usuarios del mismo. La encuesta se realiza a la primera promoción que se ha formado en el nuevo plan de estudios un vez finalizada su formación dentro del Centro Universitario de la Defensa antes de su marcha a las academias especiales. Se requiere información referente a tres ejes fundamentales: formación en el centro, prácticas internas y prácticas externas. Los resultados de la encuesta permiten afirmar que el diseño, la comunicación y la implementación del currículo es evaluado positivamente por los cadetes. Sin embargo, las acciones de capacitación no curricular, como son las prácticas externas e internas y el desarrollo de actividades de liderazgo en situaciones reales, que se consideran complementarias a la educación formal, reciben una evaluación baja, considerándolas no relacionadas con el liderazgo. Esta evidencia sugiere aspectos a ser revisados en el nuevo sistema educativo con el objetivo de mejorar su efectividad.

Palabras clave: educación, liderazgo, plan de estudios, transformación, recursos humanos

Introduction

Defense education has seen remarkable expansion and public attention in the last twenty-five years. The training apparatus that prepares armed forces is now considered as an important educational system for the nation and its security (Barret, 2009). This interest has come along with the evolution of military training and education. There are several reasons to this evolution. First, societies, whose armed forces protect, have changed (Libel, 2016). Second, although the main mission of armed forces continues to be ensuring national defense some additional functions have emerged: peace support operations, help in natural disasters and fighting terrorism. The mix of these traditional and new functions has changed (Caforio, 2007). In third place, the evolution of international security landscape has had significant implications for military professionalism and multinational military relations and alliances (Williams, 2008). Two valuable examples can be acknowledged; namely the North Atlantic Treaty Organization (NATO) enlargement process and the European Common Security and Defense Policy (CSPD). Both processes, from a sociological point of view, are even more interesting since they evidence the development of the military epistemic community and its influence; "military culture and expertise have spread internationally along time and it is socially recognized" (Cross, 2011 p.152). As new challenges arrive, both NATO and CSPD show that "training and education is one of the main reasons for its continuity and increasing leadership" (Cross, 2011 p.27).

All these factors seem to require a different military preparation than the one that used to characterize Western armies in the past (Caforio, 2007). More than ever before, not only technical knowledge but also social skills are principal to acquire the competencies army officers need: leadership, "cultural interoperability", highly flexible organization and mentally agility to size up complex situations rapidly (Boëne, 2008). Enhancing leadership skills of military leaders is of primary importance to the Army (Steele & Garven, 2009). Military leadership is stressed and developed through several mechanisms including formal education, operational assignments and self-development (Wong et al., 2003). This process also reinforces the military culture convergence and is a precondition for robust European security cooperation (King, 2006). Accordingly, interdisciplinary social courses and programs such as

Received: 10-05-2018 Accepted: 29-03-2019

"leadership", "social psychology", "organizational behavior" and "group dynamics" have entered military academic curricula in Western countries (Segal & Ender, 2008).

Although Spain is a NATO and EU member since 1982 and 1986 respectively, it did not initiate a comprehensive reform of the military training and education system until 2007.[2] This reform improves military training and education in two respects. On the one hand, military education adjusts to the new parameters of the European Higher Education Area developed by the Bologna process. On the other hand, Spain's reform stresses the increasing value of education through better leadership, decision-making and more efficient communications systems' use.

The objective of this paper is to study the recent evolution of leadership development in the Spanish Military education system. We consider this question interesting because in the current changing strategic scenario, armies need to develop a new kind of leaders: adaptive leaders, experts in the art and science of the profession of arms who are innovative, flexible and able to quickly adapt to the wide-ranging conditions (U.S. Department of the Army, 2008a). Furthermore, exhaustive management of education that allows the development of its human resources is one of the most challenging human resources (HR) strategic actions an organization can plan (Hedge et al., 2006). This is particularly relevant for military organizations that educate their members themselves from the start and who are central to their mission accomplishment.

For that, we use the framework of Wright and Nishii (2006), a model where HR practices spans in multiple levels of analysis: from the intended and actual practices to the individual perceptions and reactions. Accordingly, we firstly analyze the intentions of the leadership focus in the academic curriculum. Second, we study the establishment of a comprehensive ambitious leadership plan as part of the communication strategy. Thirdly, we turn to the perceptions of the cadets about the new system. Finally, based on the survey results, we propose some amendments to the system implementation.

The remainder of the paper is organized as follows. We first describe the Spanish General Military Academy. Second, we analyze the intended/actual educational program designed to develop the talented young

(2) Law 39/2007

Received: 10-05-2018 Accepted: 29-03-2019

women and men enrolled, paying special attention to the new leadership course and activities introduced. We then describe the communication process and efforts to integrate different cadets' activities. Afterwards, we present the cadets' reactions to the new program. Finally, we summarize the importance of relating education systems to leadership development and conclude.

Spanish General Military Academy

The Spanish General Military Academy (GMA) was founded in 1882 and throughout three discontinued periods of activity, has trained more than 23,000 officers in the army, civil guard and tri-service corps members of the Spanish armed forces. Likewise Western military academies the traditional educational model has been based on mathematical sciences and military courses with limited space to humanistic area. Although the Spanish military career has adapted to the changing needs of the Army and the arrival of new technologies, the percentage of humanities courses has never exceeded 20% (below 10% previous to the last reform (Izquierdo et al., 2011). This low percentage contrasts with the general trend to increase humanistic courses in military education (Caforio, 2007).

Discipline and values are the cornerstone of the education system. Since learning occurs though observation as highlighted by the *social learning theories* (Bandura, 1977), cadets are prompted to lead by example and follow the behavior of the best. The moral values rely on the Ten Cadet Commandaments (Cadet Honour Code) and the Royal Ordinances[3] of the Spanish armed forces (Ministry of Defence, 2009). The Cadet Honour Code is a formal statement of the honorable conduct expected by cadets and was first introduced in 1927. It has maintained with little changes. The Royal Ordinances collect the ethical principles that must guide military profession. They have also been updated and modernized to stay abreast of political and social changes. These codes together with *spirit of the General,* which emphasizes comradeship among its members built up through the close coexistence of all the students who undergo a hard daily routine, constitute the very moral guidelines of cadets.

(3) Royal Decree 96/2009.

Received: 10-05-2018 Accepted: 29-03-2019

Implementation of the New Education System

Long-term defense reform starts with defense education (Barrett, 2009). The new education model belongs to the new European-wide concept of military education, which combines vocational training and academic education ending up in a university degree (Paile, 2010).[4] The central role played by the GMA is completed by the civilian local University of Zaragoza through a partnership with the University Centre of Defense (UCD). Cadets study four years at GMA. During the fifth year, cadets shadow a military unit external to the GMA and attend specific military branch academies. The curriculum is designed taking into account what the cadet/student is expected to know at the end of the degree program: both generic skills and specific knowledge (Kehm, 2010).

The adopted curriculum reflects both the strategy defined in the 2004 national defense directive and the influence of the European and NATO military epistemic community. This decision intends first to better prepare armed forces for future challenges. Second to enhance academic standards and to deepen the professionalism of military education and finally increase international collaboration. Whatever future conflicts look like they will be tackled in cooperation with partners and allies. Therefore, it is not so important to be good at everything but rather to become more compatible and complementary (Sookermany, 2011). For that it is crucial to have peer external recognition and increase institutional relationships between military academies.

Leadership development is central in the new Spanish defense education. Military leadership has been traditionally associated to the role of chief more than that of leader. The two are tightly bound together, but a chief is not necessarily a leader in, in the current perspective of leadership (Robbins & Judge, 2013, Yukl, 2010). Therefore, army officer leadership training should focus on the training of competencies and skills to lead, common to strategic organizational management and not only to the actions of the command (chief) linked to the specific competences

(4) The main distinctive feature of the Spanish case compared to other European countries is that the General Military Academy status is not transformed to University nor a military university is established either. Instead, the academic curriculum is entrusted to a new-born university centre affiliated to a civil university. Despite this different status, 82% of the European military education institutions are recognised as Higher Education Institutions (HEI) (Callado Muñoz & Utrero González, 2016).

of his military training. This double approach is pursued in the cadet education through: (1) formal leadership course (technical, theoretical, personal and social competencies) contextualized in the military sphere, and (2) a leadership plan (complementary training not curricular). The training of the leader is posed as a specific role (competences), in equality between genres during training and instruction process, as well as the requirements of the military career (DeRue et al., 2011). Other issue is the access of women to armed forces, which entails contextualizing the role of women in the Postmodern Society and the societat struggle at the technical, historical and, not only, cultural level (Swain & Koreman, 2018).

This shift shows that Spanish armed forces are aware that new challenges need a new kind of leadership more psychology–oriented based (Segal & Insoo, 2014). The inclusion of the academic leadership course is in line with the educational fields to be improved suggested by Caforio (2007) and Nuciari (2007) and it is similar to the approach taken in the US Military Academies, Swedish National Defence College and Portuguese General Academy among others. Further, military courses are redesigned to include military values and leadership from the first year to develop an understanding of army leadership that facilitate maximum leadership effectiveness (Hatfield, 1997). Moreover, the leadership course has been established in the second year. Accordingly, cadets acquire assessable and trainable competences and skills that can be exercised so forth. Self-development of leadership skills is a critical step in the process of developing Army leaders (Flowers, 2004).

Extra-curricular activities conform the holistic leadership plan to integrate and reinforce all the associated tasks and evaluate the results.

The bachelor program

A special management engineering degree adapted to the specific competences army officers should acquire is implemented. It integrates maths, engineering, business administration, leadership and human resources. Although, each military branch has its own specific program leadership course is common to all branches.

Received: 10-05-2018 Accepted: 29-03-2019

Leadership course

Leadership is studied as a group phenomenon: the leader influences organizational management in a strategic, effective and adaptive manner. She/he knows that reality is dynamic and changing, and this implies knowing what style of leadership should be exercised and when to do it to reach the objectives. Each leadership style implies applying certain cognitive, behavioral and emotional skills for the benefit of the team. The question is what the subordinates do require from me to facilitate objective accomplishment. Cadets must acknowledge what and how to apply the best leadership style according to the task, objective, organizational context and the evolution of the group that leads at each moment.

This approach implies equipping students with intra-interpersonal and group competencies for the management of teams and work groups, without forgetting the self-leadership training inherent in leadership and leadership actions (Boëne, 2008, Caforio, 2007, Cross, 2011). The leader requires personal knowledge to lead (Goleman, 2011), and technical knowledge about people, groups, organizations and societies to guide and influence the achievement of objectives (Gil and Alcover, 2005). From this approach, leadership course has six essential axes. A first level of training in self-leadership, focused on four basic areas: (a) self-awareness, (b) self-regulation, (c) social awareness and (d) management of relationships (Goleman, 2013). The objective is to increase the leader's introspection in their strengths and weaknesses, and reach their level of excellence after each of their actions, activating the processes of cognitive-emotional revaluation.

A second level implies training the leader to manage the team as a technician: evaluate, analyze and diagnose the structural and process criteria of the group to manage its evolution. The leader should apply social and appropriate skills for the group benefit. The central element of leadership action is to understand the group as a unit of analysis, integrated and dynamic in time and space that requires specific group actions (interaction, interdependence and exchange) to articulate its operation and exert influence on it. It also adds a broad and multi-theoretical knowledge of leadership styles to convey the idea that there is not a single useful style.

Received: 10-05-2018 Accepted: 29-03-2019

A third level refers to understanding the psychosocial aspects of social interaction in the development of leadership. The perceptual aspects of the interpersonal interaction that any leader must manage to generate influence in people. It is about knowing socio-cognitive phenomena inherent in social interactions to exert influences (Halo Effect, Egocentrism, Heuristic, Pygmalion Effect, prejudices, formation and change of attitudes, etc.).

The fourth level works on the skills of management of work teams and communication. Special emphasis is paid to the managerial attitude in work groups communication and to recognizing dysfunctional behaviors that hinder the communicative process.

The fifth level provides cadets with strategies to deal with high-risk situations both from an intrapersonal perspective and for their teams. The approach is from the Positive Organizational Psychology, and socializes the cadet to deal with crisis situations: identifies dysfunctional aspects in the team and in people subjected to situations of high emotional impact.

Finally, the sixth level focuses on leadership and society. In the military career, the officer must assume different national and international destinations; carry out humanitarian actions in times of social convulsion of war and peace (security and defense). The leader must manage his team in situations of disturbances, wars and natural disasters, as well as mass immigration processes for humanitarian reasons. The officer must have certain competences for the management of collective action in situations of emergencies and crises, where social contagion can trigger chaos and must acquire notions of psychosocial logistics to manage the mass and the multitude. In this module, the army is also qualified to direct community social action projects.

Leadership Plan and Communication of the New System

The leadership plan of the GMA completes the leadership training. It is coordinated by the military although strong cooperation with the academic education is enforced. It builds on the long experience of GMA but adapted to the new challenges army face. It is based on the tenants of transformational leadership (Bass, 1998), which essentially encourages cadets to develop and respect own values and ethical perspectives thereby teaching how to become leaders in their own right without

coercing followers (MADOC-DIDOM, 2007). The plan is cross-curricular and implemented from the very first training year to reinforce formal education and training.

In terms of the new education system analysis, this plan shows the organization efforts to communicate and show the relevance of leadership to cadets (Wright & Nishii, 2006). In particular, it integrates all the activities, although apparently unrelated, to leadership, to highlight that as U.S. Department of the Army (2008b, 4-8) states "leadership can be the difference between success and failure". In this sense, developing leadership skills should be incorporated at all levels of military training (Flowers, 2004) since putting into practice the values learnt from the very start constitutes a continuous and accumulative learning system, as cadets learn from mistakes and mimic model behavior. The underlying message is that leaders, who are able to have great effects on followers by appealing to their ideals and values, can produce a higher level of commitment to the vision and mission of the organization (Piccolo & Colquitt 2006). Furthermore, cadets are provided with individual guidance by professors-tutors within group and person-to-person meetings in order to emphasize the synergies between all activities, control cadet evolution and propose and suggest additional effort when necessary for the effective learning.

The three pillars of the Leadership plan are competencies, capabilities and values and is organized around four different types of activities. Table I collects the activities programming along the five years.

TABLE I. Educational activities of the Leadership plan along the academic period

Educational activities	Year				
	1	2	3	4	5
Case Study/ Formation in Values/Military Curricular Activities	x	x	x		
Leadership Seminars			x		x
Leadership practices	x	x	x	x	x
Complementary Activities	x	x	x	x	

Note: x means that the activity is carried out during the corresponding year of the degree.

Received: 10-05-2018 Accepted: 29-03-2019

First, additional educational activities are designed to complement the leadership course: i) case studies and value analysis are set to analyze how values can influence military leadership, ii) leadership seminars are carried out that complement the structured seminars in the leadership course. These allow students promoting and encouraging group discussion, putting into value the theoretical knowledge acquired to understand real cases. Furthermore, an annual conference cycle, *Cervantes Chair*, is scheduled to reinforce and develop cultural and humanistic values to complete the comprehensive cadet education.

Cadets (from the first year) have different operational assignments according to Army rules and GMA statutes to complete their leadership training. For instance, they are assigned security tasks (with different responsibility according to the year) or head of class and room. In this sense, heads of class and room imply putting into practice the acquired competences, capabilities and values since they allow cadets to have the main responsibility of its section (class) or room during one week. They include interaction with captains on a day-by-day basis, with teachers within class schedule and any other additional task related to the management and guidance of the group.

Moreover, from the third year cadets have specific command assignments. Third year cadets hold leadership positions becoming auxiliary commands for the newcomers' three week military instruction. Similarly, fourth year cadets become auxiliary commands for the military instruction and training activities of the first and second year cadets. There are also commissioned theoretical lessons and practices as well as different adjunct chief positions (section, company or battalion) to help drill instructors. Furthermore, they are in charge of battalions during the weekly flag raising ceremony. During the fifth year, cadets shadow a unit of their military branch for six weeks. During this period they can put into practice what strategic leadership management implies, and apply the trained technical, personal and social skills acquired with the leadership course.

Apart from these structured activities, cadets participate in other activities to reinforce the leadership training. The *pater program* (mentorship plan) is the main one. It consists of students from second year taking first year mates and help them in their process of adaptation and integration in the GMA in all aspects. Sometimes it becomes a relationship that goes beyond the stay at GMA and accompanies former cadets over

Received: 10-05-2018 Accepted: 29-03-2019

their military life. Recently the traditional ceremony of *presentation of sabres* has been connected to the mentorship program (the second year cadet delivers the saver to his mentee) to reinforce the bonds developed. In this sense the second year cadet earn not only respect (through the *presentation of sabres*) but also trust from the newcomer; turning into a leader for her/his. Although the application of principles and values turn to be much more complex when they involve people they study and live with, it helps to show the potential for organizing team activities and developing leadership, especially provided roles are assigned, understood and assessed in a comprehensive manner (Cycyota et al., 2011).

Finally, the leadership plan includes a 360 degrees evaluation that refers to multi-source feedback that originates from superiors, subordinates, peers, and self-rates. The purpose is to gain insights on one's own leader behavior from others who provide valid ratings (Karrasch et al., 1997) and has proven useful since feedback from multiple sources can facilitate leader development (Steele & Garven, 2009).

Feedback on the New System

When considering the success of any organization initiative, it is important to look at both individual and group reactions. Experiences in other activities or past activities will influence the way cadets perceive and evaluate the whole leadership development process (individual variance). The involvement of good instructors and professors are key to increase success since they can reduce this variance (Wright & Nishii, 2006). Furthermore, increased individual performance may depend upon the kind of interdependence that exists among the group (Kozlowski & Klein, 2000). Whereas group norms regarding appropriate behavior may impede the expected results (Barsade & Gibson, 1998), a shared understanding of the task, team and situation affect the way the group coordinate their actions to the achievement of positive outcomes (Cannon-Bowers et al., 1993).

In order to acknowledge cadets' perceptions and interrelations, a survey was conducted in October 2017. The purpose is to evaluate how far leadership training and development has been appropriate to the current military education system. This approach does not search for the best practice, but instead seeks to assess whether cadets consider

Received: 10-05-2018 Accepted: 29-03-2019

they were adequately prepared for their current roles and that there were no relevant individual variance and/or negative implications from group interrelations. The survey consists of fifteen questions structured in three parts, questions on (1) the leadership education received, (2) the leadership practices and activities realized in the GMA and (3) the shadowing a military unit. A Likert scale where 1 is completely disagree and 8 is completely agreement was used. The reason to choose this scale is because this kind of scales with labels attached to each point on the scale have been extensively tested in social science literature (Garland, 1991). Furthermore, the cadets are invited to highlight which aspects should be improved and make suggestions accordingly in an open answer question. The questionnaire was distributed electronically to the 145 cadets enrolled in the fifth year in the 2017-2018 academic year. The survey was launched after the shadow of a unit period, during the week staying at GMA and before leaving for the branch academies. A total of 95 questionnaires were returned with a response rate of 66%. As we are interested in individual variability, we analyze not only the average answer but also the standard deviation to each question. Further, we investigate the potential group variance taking into account the cadet respondent branch.

Results and discussion

Table II shows the results of cadet's evaluations of his leadership training as officers in the Army during the four years at the GMA. The training obtained in the GMA (M = 6.03, SD = 1.31) is above the one performed in the shadowing of a unit (M = 5.69, DT = 1.67), and in the leadership position (M = 5.51, DT = 1.85). The most homogeneous distribution corresponds to the first block (formation in the GMA) due to the socialization process during four years, where they are mostly in boarding regime, and with a very firm, standardized regulation and norm (military and civil culture). The maximum dispersion is obtained in the second (leadership position) and third block (shadowing a unit). Both leadership positions and shadowing a unit are subject to process variability, since the activities to be carried out are common, but the procedure of these, how they are carried out, depend on technical variables (structural and process) linked to the equipment, relationship with superiors, and destination unit. Note that the best valued leadership

Received: 10-05-2018 Accepted: 29-03-2019

practices are shadowing of a unit compared to leadership positions (performed in the GMA), as can be seen in Table II.

TABLE II. Questionnaire results

GMA	Mean	Std. Dev.	Significance
1. Does the training obtained during the four years at GMA enable me to lead as a Lieutenant?	6.02	1.34	2*, 3***, 4**, 5**
2. I have enough knowledge to be a leader today	6.21	1.13	3***, 4***
3. I have the technical knowledge (skills, competencies and procedures) necessary to exercise leadership as a future Lieutenant	5.76	1.24	5***, 6**
4. I have the theoretical knowledge (concepts) necessary to exercise leadership as a future	5.74	1.43	5***, 6**
5. They have given me enough values to exercise leadership as a future Lieutenant	6.32	1.45	
6. I was able to carry out tasks of leader and chief during the period of external practices.	6.14	1.64	
GMA Cadet Leadership Position	**Mean**	**Std. Dev.**	**Significance**
7. The training on leadership obtained at GMA has allowed me to gain experience to lead the subordinates	5.68	1.64	8***, 9***, 10***, 11***
8. The practice carried out with the first year students has served me in my training as a future Lieutenant	6.15	1.78	9***, 10***, 11***
9. Developing the figure of Pater with the newcomer cadets at the GMA has helped me to develop leadership functions as a future Lieutenant	3.96	2.09	
10. Having a partner of superior courses like Pater has helped me in my training as a cadet at the GMA	3.86	2.25	
11. The activity of the "Cervantes Chair" has given me useful knowledge for my global training as a future Lieutenant of the army.	3.88	2.20	
Shadow a Unit	**Mean**	**Std. Dev.**	**Significance**
12. I have exercised leadership during the execution of external practices	6.06	1.68	13*, 14***, 15***
13. I have observed differences between the role of leader and chief in the development of functions of an Army officer	6.34	1.62	14***, 15***
14. I have observed that the officer performs only leadership functions in leadership management	4.94	2.16	15***
15. I have observed that the officer performs only chief tasks in leadership management.	4.26	2.13	

Legend: numbers of the significance column indicate significant differences in means with the corresponding question and *, **, *** the level of significance: 0.1, 0.05 and 0.01 respectively.

Received: 10-05-2018 Accepted: 29-03-2019

The cadets appreciate, in a very satisfactory and significant way, the curricular training obtained from leadership in the GMA. When asked about the knowledge received, values stand out against technical and theoretical aspects, and clearly differentiate their role as chiefs and leaders in leadership. There is little difference between the scores obtained in this first block, but the data distribution indicates less dispersion in the assessment of values and knowledge compared to the technical knowledge (skills, competences and procedures) obtained in the GMA. The question on values shows a pattern of behavior different from the rest, with more atypical results (see Table II and Figure A.1 of the appendix). These data confirm the vision of the implementation of the new curriculum in the GMA, not only to transmit concepts, but to include technical skills and have an outstanding impact on the transmission of values in a comprehensive training of the future officer as a military leader (Caforio, 2007; Hatfield, 1997; Kehm 2010)

The response pattern on leadership positions in the GMA (second block) focused on leadership presents more dispersion in the data than that obtained in the first block (training). The cadets appraise in a more homogeneous way the received training (transmission of concepts, procedures and values) than the internal practices carried out in the GMA in their leadership training. These practices are valued in a positive and meaningful way when they lead the subordinates in the command action (GMA7). The activities that achieve the lowest, most dispersed scores are those oriented to coaching (Pater figure, GMA9 and those made with a superior classmate, GMA10), and training (attendance at the *Cervantes Chair*; GMA11). These two have the greatest dispersion in the response pattern (See Table II and Figure A.2 of the appendix). These data indicate that the practices linked to the leadership plan must be reviewed in a significant way so that they are meaningful for the cadets, and help integrate the theoretical into the practical. The practices directly related to the command action, where they can integrate leadership competencies are better valued (GMA12). The cadet has a technical and precise mental structure of applied practices focused on the strategic management of leadership (GMA2). This approach has already been introduced in other military agencies as explained by Smith and Bergin (2012), developing the concepts of content, procedures and values in formal education, with operational tasks and self-development activities (Wong, et al., 2003). The practical activities must be contextualized in the strategic management of leadership, as this is key in the integral formation of the leader. Otherwise they lose vision and efficacy.

Received: 10-05-2018 Accepted: 29-03-2019

The data of the third block (shadowing of a unit) show that the cadets have been able to exercise leadership during the practices, differentiating in their command the role of leader and chief with significant values (Table II), although there is more dispersion in the response pattern of these last two items (Figure A.3 in the appendix). In the formation of the military leader, it is essential to highlight the dual role (leader and chief) in leadership management in the competitive education of the future army officer These data highlight how the implementation of the new curriculum is influencing the integration of the technical competencies of the leader (management) in the military context (command action), developing a comprehensive vision of the resource manager.

Cadet claims from the open answer question are in line with the evidence presented. They suggest increasing academic courses on leadership and human resources and propose to reform non curricular activities, such as *pater program* and *Cervantes Chair* to increase its usefulness.

The response pattern varies according to the branch to which the cadets belong (table III). Infantry is the one that obtains higher scores in the first block (formation in the GMA on leadership), and Engineering the lowest, being the difference significant. The item on the transmission of values to exercise leadership (GMA5) obtains in all branches an average score of 6, being the most valued item in the entire block. This result is consistent with the strategy of including military values and leadership for a greater understanding of the army officer (Hatfield, 1997) within the implementation of the new curriculum in order to achieve maximum officer efficiency. It should be noted that the Engineering branch continues to have a downward trend in leadership training, except for item GMA6 (exercising leadership tasks) that obtains the highest score (M = 6.69). Engineering and Infantry branches have inverse patterns in their assessment, the former emphasizes theoretical and technical training, the latter practice. In the second block (leadership positions at GMA) the transmission branch obtains the highest average rating. Infantry instead stands out in practices with first-year students (GMA8), and Cavalry presents the lowest linked to the activity of the *Cervantes* chair (GMA11).

In the third block (shadowing a unit) Cavalry has the highest values whereas artillery presents the lowest. Infantry stands out for valuing the difference between the role of leader and chief in a more prominent way (GMA 13, see Table III A and B). These data show different assessment profiles according to the branch. Each branch has different competencies that affect their leadership role in a distinct way. Branches with more abstract competences (engineering) value more the practical training

Received: 10-05-2018 Accepted: 29-03-2019

meanwhile Infantry (management skills) consider more important the theoretical and technical training of leadership. On the contrary, transmissions and Cavalry value the applied versus theoretical training. Artillery has a medium and more adaptive profile for not emphasizing polarities in its response pattern.

TABLE III. Branch questionnaire results

	Panel A: Mean by branch					Panel B: Significance			
	Art	Cav	Inf	Eng	Tran	Art	Cav	Inf	Eng
1. Does the training obtained during the four years at GMA enable me to lead as a Lieutenant?	5.91	6.08	6.71	5.23	6.10	I***, E*	I**, E**	E***, T*	T**
2. I have enough knowledge to be a leader today	6.09	6.42	6.47	5.77	6.35		E*	E*	T*
3. I have the technical knowledge (skills, competencies and procedures) necessary to exercise leadership as a future Lieutenant	5.66	5.75	6.29	5.08	5.90	I*	I*, E*	E**	T**
4. I have the theoretical knowledge (concepts) necessary to exercise leadership as a future	5.56	5.75	6.24	5.15	6.00	I*		E**	T*
5. They have given me enough values to exercise leadership as a future Lieutenant	6.22	6.58	6.76	5.46	6.50		E*	E**	T*
6. I was able to carry out tasks of leader and chief during the period of external practices.	5.97	6.33	5.94	6.69	6.10	E*		E*	
7. The training on leadership obtained at GMA has allowed me to gain experience to lead the subordinates	5.63	6.17	5.88	4.92	5.80		E*	E*	T*
8. The practice carried out with the first year students has served me in my training as a future Lieutenant	6.22	6.58	6.29	5.54	6.05		E*		
9. Developing the figure of Pater with the newcomer cadets at the GMA has helped me to develop leadership functions as a future Lieutenant	3.84	3.83	4.00	3.77	4.30				
10. Having a partner of superior courses like Pater has helped me in my training as a cadet at the GMA	3.94	3.92	3.65	3.46	4.15				
11. The activity of the "Cervantes Chair" has given me useful knowledge for my global training as a future Lieutenant of the army.	3.66	3.42	4.00	4.00	4.35		T*		
12. I have exercised leadership during the execution of external practices	5.75	6.58	6.18	6.23	6.05	C**			
13. I have observed differences between the role of leader and chief in the development of functions of an Army officer	6.16	6.75	6.76	5.54	6.55	C*	E**	E**	T*
14. I have observed that the officer performs only leadership functions in leadership management	4.72	5.42	5.24	4.69	4.90				
15. I have observed that the officer performs only chief tasks in leadership management.	4.50	4.00	4.18	3.77	4.40				

Legend: Art, Cav, Inf, Eng, Tran stand for Artillery, Cavalry, Infantry, Engineers and Transmission respectively. Panel A: displays sample means. Panel B: letters indicate significant differences in means with the corresponding branch and *, **, *** the level of significance: 0.1, 0.05 and 0.01 respectively.

Concluding Remarks

The new educational system in the Spanish armed forces is now part of the European system of higher education and provides a new vision of the officer's training. The challenge has been to integrate a technical profile and humanistic training that affects the officer's organizational management abilities. Leadership education involves training in the strategic management of organizational leadership, integrating their role as leader and chief. In the new educational system, a leader must be trained in specific competences about knowledge, know how to do, how to think and how to be or feel. Self-leadership is encouraged to look for the best leader in order to obtain the best possible team, thanks to a process of continuous training of their abilities, aptitudes, values and a continuous process of introspection (Goleman, 2013). The objective is to facilitate excellence in management skills for the officer as leader himself and with his units. The role of the leader has been exposed as an expert analyst (technician) with social skills, capable of influencing people, groups, organizations and at the social level. The implementation of the new curriculum integrates these assumptions and obtains a positive evaluation by cadets.

The evaluation of the cadets is more favorable with respect to the training received in leadership at the curricular level (concepts, procedures and values), than the non-curricular training actions (internal practices) in the GMA. These actions designed in the first place to facilitate the process of socialization of the newcomer cadet on organizational culture (formal and informal) receive lower assessment by the cadets as they are perceived unrelated and not useful. This evidence suggests aspects to be reviewed in the GMA leadership plan and communication with the objective of improving its effectiveness. In particular, non-curricular activities aimed to strengthen the leadership acquired at the curricular level. This assessment is in line with the cadets requesting more structured leadership practices together with additional academic courses on leadership and human resources management. Moreover, cadets suggest implementing courses with actual real cases as well as organizing seminars with expert professionals.

The values transmitted in the GMA are assessed satisfactorily but results indicate heterogeneity in the response pattern. This reflects the lack of homogeneity in the transmission of these values and suggests

Received: 10-05-2018 Accepted: 29-03-2019

the revision of channels and activities oriented towards this objective. In addition, results suggest that applied training must go through activities directly linked to the strategic management of leadership. They should focus on the importance of generating influence in subordinates during the action of command, integrating the role of leader and chief in the strategic management of leadership. These activities must have a clear and direct transfer to integrate theoretical and technical concepts into practice.

A final question refers to the possibility of incorporating a more adapted curriculum by branch. In the case of leadership it may be necessary to review if the strategic management of leadership should contemplate a generic formation (like the current one) and/or a specific one to adapt the management of the command to the different branches.

As suggested by Kirchner &Akdere (2017) for the U.S., this comprehensive approach to leadership training and lessons learnt by the Spanish army could be potentially useful for non-military organizations interested in starting leadership training early in the employees' career including all employees.

References

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ, Prentice Hall.

Barrett, J. (2009). Modern developments in Defense education. *Connections* 8(2), 1-5.

Barsade, S.G., & Gibson, D. E. (1998). Group emotion, a view from top and bottom. In D. H. Gruenfeld (Ed), *Composition. Research on managing groups and teams* (vol. 1, pp. 81-102). Stamford, CT. JAI Press.

Bass, B. M. (1998). *Transformational leadership, Industrial, military, and educational impact*. New Jersey, Erlbaum.

Boëne, B. (2008). Teaching sociology at Saint-Cyr, 1983-2004 and beyond. A personal account. *Armed Forces & Society*, 35(1), 16-35.

Received: 10-05-2018 Accepted: 29-03-2019

Caforio, G. (2007). Officers' professional preparation and issues posed by new military missions. *International Review of Sociology*, 17(1), 87-104.

Callado Muñoz, F.J. & Utrero-González, N. (2016) Educación Superior en Defensa y Seguridad: Una Perspectiva Europea. Ed. José Serna Serrano, Pilar Sánchez-Andrada y Ignacio Álvarez Rodríguez *Libro actas DESEi+d 2016*, San Javier.

Cannon-Bowers, J. A., Salas, E., & Converse, S. A. (1993). Shared mental models in expert decision making. In N.J. Castellan (Ed), *Individual and group decision making* (pp. 221–246). Hillsdale, NJ, LEA.

Cross, M. (2011). *Security integration in Europe*. Ann Arbor: University of Michigan Press.

Cycyota, C. S., Ferrante, C. J., Green, S. G., Heppard, K. A., & Karolick, D. M. (2011). Leaders of character, the USAFA approach to ethics education and leadership development. *Journal of Academic Ethics,* 9, 177-192.

DeRue, D. S., Sitkin, , S. B., & Podolny, J. M. (2011). From the Guest Editors: Teaching Leadership-Issues and Insights. Academy of Management. *Learning & Education, 10 (3),* 369-372. http://dx.doi.org.sire.ub.edu/10.5465/amle.2011.0004

Flowers, M. (2004). Improving strategic leadership. *Military Review*, March-April, 40-46.

Garland, R. (1991). The Mid-Point on a Rating Scale, Is it Desirable?. *Marketing Bulletin*, 2, 66-70.

Gil, F., & Alcover, C. M. (2005). *Introducción a la psicología de los grupos*. Madrid: Pirámide.

Goleman, G. (2013). *Liderazgo. El poder de la inteligencia emocional*. Barcelona: Ediciones B

Hatfield, B. Jr. (1997). Strategic leadership development: An operation domain application. Research Paper (0607). Maxwell-Gunter Air Force Base, Montgomery, Alabama, Air Command and Staff College.

Hedge, J.W., Borman, W.C., & Bourne, M. J. (2006). Designing a system for career development and advancement in the U.S. Navy. *Human Resource Management Review*, 16, 340–355.

Izquierdo, J., Ortíz de Zárate, J. R., & Aparicio, A. (2012). *La Academia General Militar. Crisol de la oficialidad española*. Zaragoza: Institución Fernando el Católico (CSIC).

Karrasch, A. I., Halpin, S. M., & Keene, S. D. (1997). Multirater assessment process-a literature review. ARI Technical Report 1076. Ft. Leavenworth, Kansas, Army Research Institute.

Kehm, B. M. (2010). Quality in European higher education. The influence of the Bologna process. *Change*, 42(3), 40-46.

King, A. (2006). Towards a European military culture?. *Defense Studies*, 6(3), 257–277.

Kirchner, M. & Akdere, M. (2017) Military leadership development strategies: implications for training in non-military organizations, *Industrial and Commercial Training*, 49(7/8), 357-364, https://doi.org/10.1108/ICT-06-2017-0047

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations, contextual, temporal and emergent processes. In K.J. Klein & S. W. J. Kozlowski, (Eds), *Multilevel theory, research and methods in organizations, Foundations, extensions and new directions* (pp. 3-90). San Francisco, CA: Jossey-Bass, Inc.

Libel, T. (2016). *European military culture and security governance*. Oxon: Routledge.

MADOC-DIDOM. (2007). *Liderazgo. Orientaciones. (Manual para uso interno de las Fuerzas Armadas)* Granada: Ministerio de Defensa. Ejército de Tierra.

Ministry of Defence (2009). *Royal Ordinances of Spanish Armed Forces. Royal Decree 96/2009.* Available at: https://www.boe.es/boe/dias/2009/02/07/pdfs/BOE-A-2009-2074.pdf

Nuciari, N. (2007) Coping with Diversity: Military and Civilian Actors in MOOTW, International Review of Sociology: Revue Internationale de Sociologie, 17:1, 25-53, DOI:10.1080/03906700601129541

Paile, S. (2010). *The European military higher education. Stocktaking report*. Brussel: ESDC.

Piccolo, R. F., & Colquitt, J. A. (2006). Transformational leadership and job behaviors, the mediating role of core job characteristics. *Academy of Management Journal*, 49(2), 327–340.

Robbins, S., & Judge, T. (2012). Liderazgo. In S. Robbins & T. Judge (Eds.), *Comportamiento Organizacional* (pp.367-409). México: Pearson.

Segal, D. R., & Ender, M. G. (2008). Sociology in military officer education. *Armed Forces & Society*, 35(1), 3-15.

Segal, D. R., & Insoo, K. (2014). Educating Military Leaders : The US and South Korea in Comparative Perspective. *Res Militaris*, 4 (2), Summer-

Autumn/ Été-Automne. http://resmilitaris.net/ressources/10186/43/res_militaris_article_segal___kim_educating_military_leaders.pdf

Smith, H., & Bergin, A. (2012). *Educating for the profession of arms in Australia*. ASPI Special Report, 48.

Sookermany, A. McD. (2011). The embodied soldier, towards a new epistemological foundation of soldiering skills in the (post) modernized Norwegian armed forces. *Armed Forces & Society*, 37(3) 469-493.

Spain (2007). *Militar Career Law. Law 39/2007.* BOE-A-2007-19880. Available at https://www.boe.es/buscar/act.php?id=BOE-A-2007-19880

Steele, J. P., & Garven, S. (2009). *The Leader AZIMUTH Check, Factor Structure of Common Competencies.* Technical Report 1250. U. S. Army Research Institute for the Behavioral & Social Sciences.

Swain, J. & Koreman, L. (2018). In their humble opinion: How expressions of humility affect superiors' assessments of leadership potential in the US Army. ***Military Psychology****, 30(6),* 507-527. https://doi-org.sire.ub.edu/10.1080/08995605.2018.1503002

U.S. Department of the Army. (2008a). *U.S. Army training and doctrine command general. TRADOC vision*. Washington, D.C.: Department of the Army. Available at http,//www.tradoc.army.mil/about.htm

U.S. Department of the Army. (2008b). *Operations (Field Manual 3-0)*. Washington, D.C: Department of the Army.

Williams, J. A. (2008). The military and society beyond the postmodern era. *Orbis*, 52(2),199-216.

Wong, L., Bliese, P., & McGurk, D. (2003). Military leadership: A context specific review. *The Leadership Quarterly*, 14(6), 657-692.

Wright, P. M., & Nishii, L. H. (2006). Strategic HRM and organizational behavior, Integrating multiple levels of analysis (CAHRS Working Paper 06-05). Ithaca, NY, Cornell University, School of Industrial and Labor Relations, Center for Advanced Human Resource Studies.

Yukl, G. (2010). Liderazgo en las organizaciones. In G. Yukl (Ed.), *Liderazgo en las organizaciones* (pp. 54-58; 93-117). Madrid: Pearson Prentice Hall.

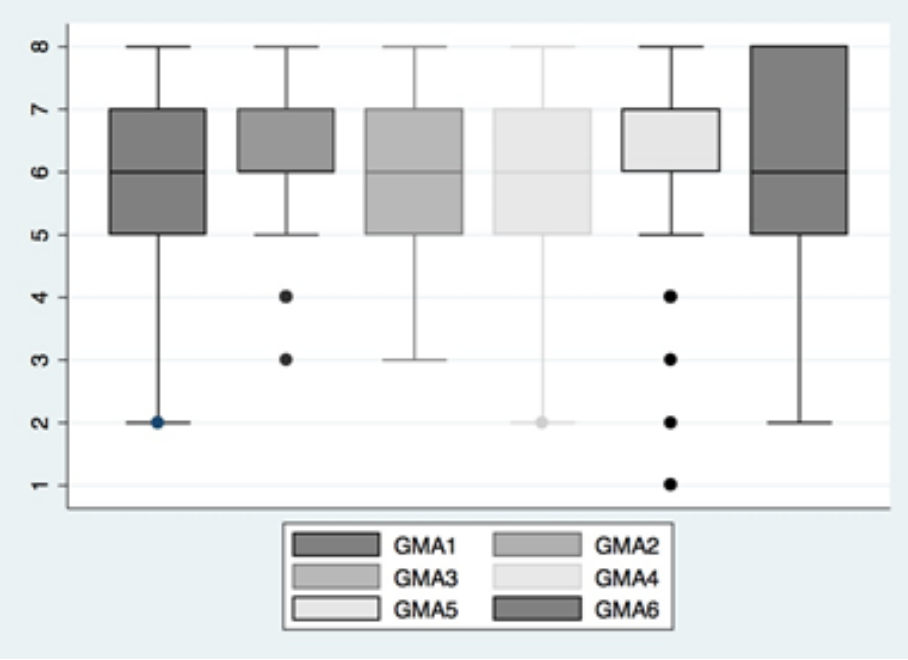
Contact address: Natalia Utrero González. Centro Universitario de la Defensa. Academia General Militar. Ctra. de Huesca s/n, 50090 Zaragoza, España. E-mail: n.utrero@unizar.es

Received: 10-05-2018 Accepted: 29-03-2019

Appendix

FIGURE I. Dispersion of results

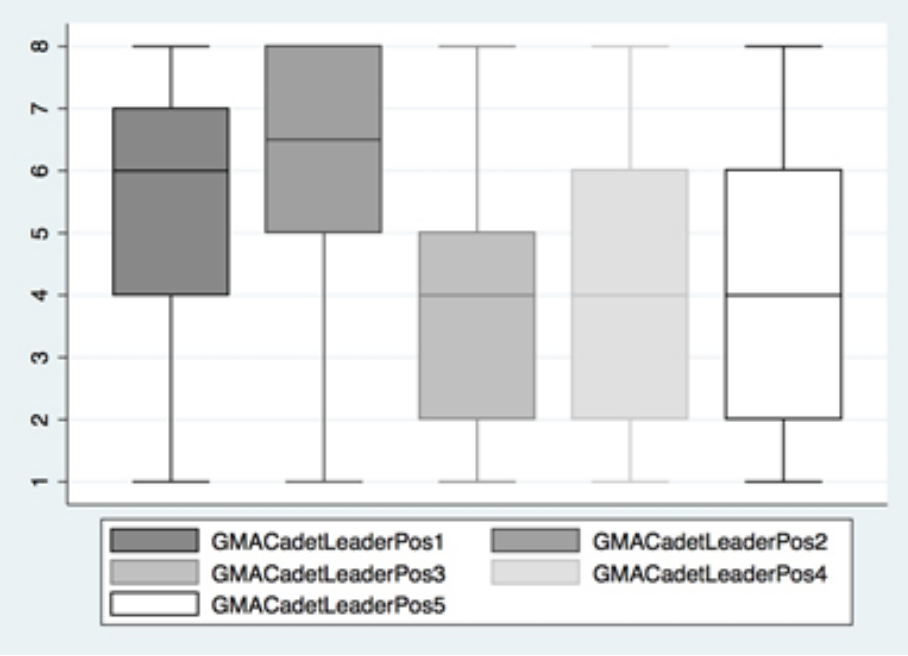
I.I. Items related to GMA formation



Legend: ***Items***: GMA1: Does the training obtained during the four years at GMA enable me to lead as a Lieutenant?; GMA2: I have enough knowledge to be a leader today; GMA3: I have the technical knowledge (skills, competencies and procedures) necessary to exercise leadership as a future Lieutenant; GMA4: I have the theoretical knowledge (concepts) necessary to exercise leadership as a future Lieutenant; GMA5: They have given me enough values to exercise leadership as a future Lieutenant; GMA6: I was able to carry out tasks of leader and chief during the period of external practices.

Received: 10-05-2018 Accepted: 29-03-2019

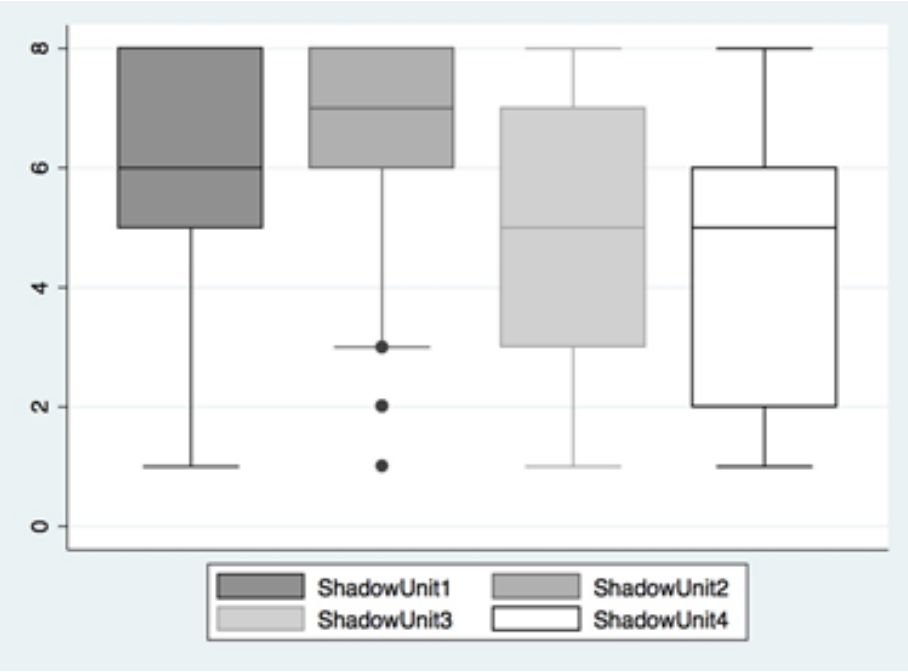
I.II. Items related to GMA cadet leadership positions



Legend: ***Items***: GMACadetLeaderPos1: The training on leadership obtained at GMA has allowed me to gain experience to lead the subordinates; GMACadetLeaderPos2: The practice carried out with the first year students has served me in my training as a future Lieutenant; GMACadetLeaderPos3: Developing the figure of Pater with the newcomer cadets at the GMA has helped me to develop leadership functions as a future Lieutenant; GMACadetLeaderPos4: Having a partner of superior courses like Pater has helped me in my training as a cadet at the GMA; GMACadetLeaderPos5: The activity of the Cervantes Chair has given me useful knowledge for my global training as a future Lieutenant of the army.

Received: 10-05-2018 Accepted: 29-03-2019

I.III. Questions related to shadowing a unit



Legend. ***Items***: ShadowUnit1: I have exercised leadership during the execution of external practices; ShadowUnit2: I have observed differences between the role of leader and chief in the development of functions of an Army officer; ShadowUnit3: I have observed that the officer performs only leadership functions in leadership management; ShadowUnit4: I have observed that the officer performs only chief tasks in leadership management.

A proposed model for systemic evaluation of basic education[1]

Propuesta de modelo de evaluación sistémica para la educación básica

DOI: 10.4438/1988-592X-RE-2019-385-420

Gabriel Álvarez-López
Javier M. Valle
Universidad Autónoma de Madrid

Abstract

Evaluation of educational systems is a reality in nearly every advanced country. In this framework, a number of different evaluation models have appeared. However, these models have already become partially or completely obsolete. This paper presents a model of systemic evaluation for basic education at the regional government level. It begins with a few introductory aspects that place the framework for the idea behind the model, its intentions, and several important characteristics. This is followed by the development of the proposal, detailing each one of its components. Next, a synthesis of the model is offered. The paper then concludes with a section on recommendations.

Key words: educational policy, external evaluation, evaluation model, education system, Autonomous Community.

Resumen

La evaluación de los sistemas educativos es una realidad presente en la práctica totalidad de los estados avanzados. En este marco, han ido apareciendo

(1) Acknowledgements: the authors wish to acknowledge the research group Recognized Research Group on "Supranational Education Policies" [Grupo de Investigación sobre Políticas Educativas Supranacionales (GIPES-UAM)]. Funding: the manuscript has been partially funded by the Ministry of Economy and Business through the project EDU2015-65743-P.

diversos modelos de evaluación. A nuestro juicio, estos modelos han quedado ya obsoletos, parcial o completamente. En este trabajo se presenta un modelo de evaluación sistémica de la educación básica a nivel autonómico. Comienza con unos aspectos introductorios que sitúan el marco en el que se inscribe la idea con la que se originó el modelo, sus intenciones y algunas características importantes. A continuación, se muestra el desarrollo de la propuesta, detallando cada uno de sus componentes. Seguidamente, se ofrece una síntesis del modelo. Y se finaliza con un apartado de recomendaciones.

Palabras clave: política educativa, evaluación externa, modelo de evaluación, sistema educativo, Comunidad Autónoma.

Introduction to the model: preliminary aspects

Spain's regional governments (called Autonomous Communities) have been assessing the education system of their basic education. However, the models they use have too many limitations. Therefore, what is proposed here is a Systemic Evaluation Model (SEM) aimed at overcoming those limitations and encompassing a more comprehensive, holistic evaluation (Álvarez – López, 2017a).

The model is expressed here graphically as a set of concentric spheres (Valle, 2008) that includes both the elements and the relationships among them. It responds to the idea of not having any hierarchies among elements in the system, that they are all equally important. If any element is not considered, the entire evaluation is irredeemably impoverished. The relationships among the elements are also multiple, circular, and seamless, since different elements can be combined, thereby making the information from the system more complex and more accurate.

Sources of inspiration for the model

The SEM draws its inspiration from sources in three areas: the results of our research, proposals from supranational organizations, and the work of relevant authors. First, the SEM is the result of work to overcome the

Received: 29-05-2018 Accepted: 12-04-2019

limitations found in the evaluations currently being carried out in the Autonomous Communities. That work analyzed the evaluation policies of the regional education system in four Autonomous Communities (Andalusia, the Canary Islands, Catalonia, and the Basque Country) selected using criteria such as evaluation experience, permanence of the evaluations, and design of programs to assess dimensions of the system beyond the students. The research followed the comparison method (García Garrido, 1996; Caballero, Manso, Matarranz and Valle, 2016), incorporating document analysis and semi-structured interviews with the political and technical personnel in charge. The results on the evaluation policies and programs revealed three aspects that the model featured here will attempt to overcome: 1) evaluation programs overly limited in their scope, methodology, and impact; 2) the prevalence of the results on the processes (methodological limitations); and 3) the yearly census-like nature of the evaluations (limitation in scope and impact). Moreover, the results showed that the evaluations in the Autonomous Communities studied did not differ much. Finally, it was found that the pathways, the investment, the autonomy of the Autonomic Evaluation Unit (AAU) and the consensus of society are key ingredients to making evaluations useful for improving the basic education system.

Secondly, the model was complemented with contributions from a supranational study (Valle, 2013) on these matters. Evaluation of education systems has been a concern and a priority on the political agenda of modern states (Calero and Choi, 2012; Eurydice, 2015) under the auspices of international organizations such as the IEA, OECD or EU. Trends such as accountability (Hanushek and Raymond, 2001) and school effectiveness (Reynolds, Chapman, Kelly, Mujis and Sammons, 2012), along with proposals such as evidence-based policies (Davies, 2004), have served to justify the need to evaluate the education system as a whole. Neither Spain as a state, nor its Autonomous Communities have been alien to these evaluating movements (Maestro, 2006; Tiana, 2014), and the SAM, as will be seen below, incorporates some of the postulates originating in supranationality.

Thirdly and lastly, this model takes into account the works of Arturo de la Orden (1997; 2009; Orden and Jornet, 2012) and Bernardo Gargallo (2003). We agree with Arturo de la Orden on three fundamental elements: 1) the need for a holistic and comprehensive perspective of any evaluation model of education systems; 2) the analysis of both

quantitative and qualitative data; and 3) the combination of internal and external evaluations. With respect to Gargallo's model, we agree with his proposal to evaluate teachers in their initial and ongoing training.

Sense and main features of the SEM

Bolívar (2016) rightly defends that evaluation is the system's tool to prove that it is guaranteeing every student's right to education. Bolívar understands the right to education as one that includes "the set of knowledge and skills necessary for the population to realize itself personally and integrate into society and the job market" (2016, p. 308). The public administration is responsible for guaranteeing this right, and evaluation of the education system is the main tool for measuring the degree of compliance with this right of which it is the guarantor.

Another important aspect is the complementarity of evaluations. It is therefore necessary to include information gathered by the schools themselves, since some aspects are not easily seen by outsiders. This complementarity must be taken in both directions, that is, the system is interested in the information collected from the internal evaluations, but the schools themselves may also make use of information from external evaluations.

The SEM is versatile enough to accommodate the different forms of evaluation that are currently being carried out, since it allows the analysis of their appropriateness, relevance, effectiveness, impact, satisfaction, performance, etc. In short, the evaluations designed can be very different and yet all of them fit into this model.

In addition, it would be beneficial for the model if other areas of knowledge were added to it, such as assessments of public policies or the evaluation of public administrations. The model allows the incorporation of these disciplines and some others.

Finally, this model could also be implemented in other Autonomous Communities. This is due to the heterogeneity of the Autonomous Communities participating in the study underlying this proposal, given that they feature geographic and demographic diversity and have a background in evaluating their regional education systems. Furthermore, it is a model that can be adapted to different educational systems, since its implementation must arise from the participation of the educational community involved. Indeed, it is a model committed to its systemic

nature and to the participation of the entire educational community in its development.

Development of the Systemic Evaluation Model (SEM)

This model makes use of General System Theory as defined by Bertalanffy (1969), because it provides an approach to conceptualizing the "educational system". According to Colom (1979), accepting Bertalanffy's contributions,

> The educational system is an open system (in terms of its relations with the social system), it is a closed system (in terms of its relations with itself), and it is a communication system (in terms of its relations with the subjects for whom it is designed) (p. 120).

Furthermore, being a system, the education system is also made up of components, that form a structure and have assigned functions that are obtained on account of the relations between them. It is therefore accepted that a system is more than the sum of its parts.

Consequently, systemic evaluation is understood as an evaluation aimed at the students, but also at the school, school administrators, teachers, educational administration and/or the different specific programs and projects. It is an evaluation that promotes decision-making about the system from all possible prisms: global, holistic, vocational, focused on improvement and, therefore, emancipatory. It is an evaluation that arises from the public administration of the Autonomous Community towards the educational system in its territory as a whole, but which maintains its political and institutional independence, with a deep sense of equity, social justice and democratic participation of the educational community.

Therefore, an evaluation of the education system is a complex task that must be understood as the culmination of a process. However, the model is theoretical, given the well-known eventualities and limitations in designing and, above all, in implementing public education policies, particularly policies referring to evaluation. It is important to clarify that the evaluation of the education system is not the sum of the evaluations of the different dimensions. A systemic perspective is necessary, one in which what is important is the diagnosis of the system, not of its

subsystems. This will have several repercussions on the model. These premises form the basis of the model proposed here.

The SEM consists of five categories (Figure I) that aim to include all possible variables for the design and implementation of a systemic evaluation program. Each of these five categories is in turn composed of different elements (Figure II) that are the true building blocks of the model.

The categories are presented below in the form of spheres. This idea of sphericity, as opposed to linear, planar, or hierarchical models, is taken from Valle (2008)'s conception of Value Spheres of Values according to which the sphere

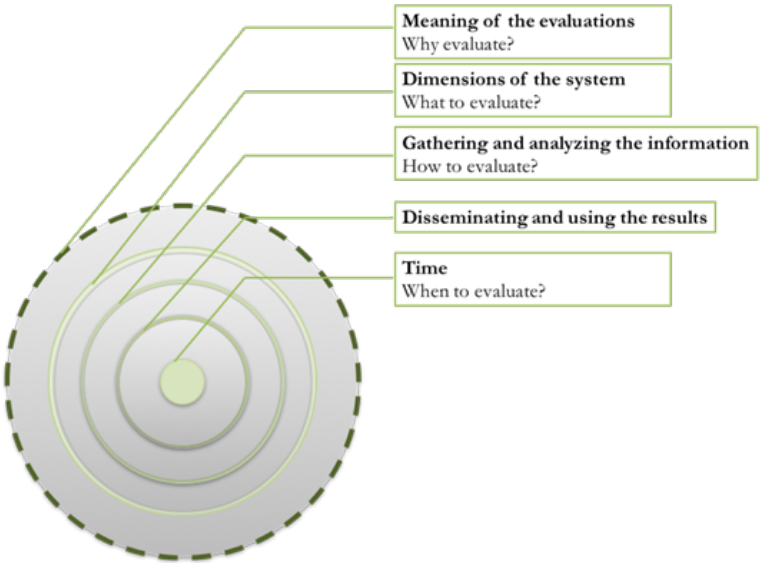
> "If we imagine it transparent, it allows different visions depending on the position from which it is looked at since our "point of view" can be altered due to structural changes (...) or in a more individual way depending on the circumstances. The sphere does not express enumeration or dominance; it expresses complex dynamic relationships" (p. 130).

The categories are presented below, from the outermost sphere to the innermost sphere. Sphere 1 (the outermost) refers to the meaning of the evaluations. Sphere 2 consists of the dimensions of the system. Sphere 3 encompasses the gathering and analysis of information. Sphere 4 refers to the results, their use, and their dissemination. Finally, Sphere 5 deals with temporality. Although the spheres have been numbered, it is not to grant any rank, but to facilitate following the explanation of the model.

All elements are equally important. If one of the elements is removed from the system, the entire system becomes decompensated.

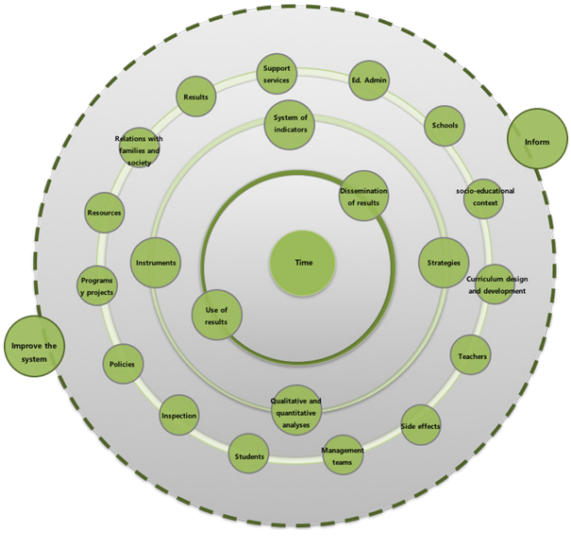
Received: 29-05-2018 Accepted: 12-04-2019

FIGURE I. Categories of the systemic evaluation model



Source: Prepared by the authors

FIGURE II. Diagram of the systemic evaluation model



Source: Prepared by the authors

Category 1. Meaning of evaluations

This category has two elements: to inform and to improve.

Discourse in which the term evaluation is used quickly reveals two underlying trends that drive it and justify it: accountability and school effectiveness. While there is no need to mistrust either one, at least initially, each does have its own concerns worth noting. Accountability, for example, may also be taken as a form of control, domination, and auditing of the components of the system. In turn, school effectiveness may be understood as forcing the logic of the marketplace onto schools, with competitive (instead of cooperative) processes and dynamics of improvement without any elements of equity.

There is an ongoing need to inform the educational community and society of what is happening in the educational system. These evaluations may help improve the processes. To inform and to improve: informing is more straightforward, but improving is full of shades of differences. The specific type of evaluation will depend on what is understood as *improvement*. System improvement may come about by analyzing aspects such as adequacy, effectiveness, suitability, impact, relevance, performance, satisfaction and/or monitoring. All these approaches can provide useful information for improving the system.

Category 2. Dimensions to be evaluated

Although the number of potential factors in an assessment of the education system may seem nearly infinite, any attempt to evaluate all of them is both unfeasible as it is irresponsible. Instead, without losing the systemic perspective, the inclusion of a wide range of system dimensions should be enough to obtain a holistic evaluation of the system. Their diagnosis and analysis will suffice to bring about overall improvement in the quality of the education system.

In the SEM proposed here, the students are not in fact the center of attention in the system evaluation. These days, the student is often the epicenter of the system: programs, teachers, schools, regional government systems, the effectiveness of the system, etc. are all evaluated through the students. Although Spain cannot be said to exert undue evaluative pressure on students, such pressure undoubtedly exists. The SEM model

Received: 29-05-2018 Accepted: 12-04-2019

diffuses that pressure. Information will be collected from the student when necessary, but also from teachers, management teams, families, educational administrators, non-teaching staff; even, if applicable, from society (for example, for an evaluation of their satisfaction with the education system).

Finally, it must be stressed that in this model all the dimensions must be evaluated in order to have an integrated, holistic view of the system. Picasso's Guernica cannot be understood by looking only at the horse. Rather, the whole picture must be analyzed in order to understand it. Similarly, the ultimate objective of these policies must lead to that global vision, unconstrained by resources, political will or context.

In short, the dimensions of the educational system taken into account here are the following: educational administration, schools, context, curriculum design and development, teachers, side effects, school management teams, students, inspection, policies, programs and projects, resources, relationship between school -family and school-society, results, and support services.

Educational administration

Evaluation of the educational administration includes analyzing the fit between workers (skills, previous training and experience) and the tasks assigned to them and the financial resources available. This evaluation also involves assessing their suitability for the post and the systems for recruiting and appointing workers and policy makers.

In addition, such an evaluation must include the organizational structure. It must determine whether the organization chart fills the needs of the policies and whether it meets the technical requirements of the evaluations. In other words, it must analyze the workers' resilience, their adaptability to change, their organizational and functional flexibility.

So far, politicians have been reluctant to subject their own educational administrations to evaluation. Evaluation of their effectiveness would involve finding out the extent to which they comply with the government program and the efficiency with which human and budgetary resources are managed.

Finally, it is also important to include how satisfied the administrators, users, and the educational community are with the quality of the service offered.

Schools

Schools are elements of the system that bring in new factors of analysis; however, the evaluation of schools here will focus only on their management and operation. Evaluation of a school requires analyzing the context (educational, social, economic and cultural, as well as the relationship the school maintains with its immediate surroundings and with its region), management of its resources (human, material, and financial), the management team (which this model considers as a dimension unto its own because of its importance), the organization and functioning of the school, the climate among students, participation (of the educational community in the school and of the school in programs and projects), and, finally, the degree of satisfaction of the educational community with regard to the school.

Context

The results obtained from the model can only be dimensioned after an evaluation of the context. Context is a concept that may too vast to cover, at the risk of evaluating aspects that do not help explain the results. The SEM proposed here offers an analysis of the context of the Autonomous Community, divided into two parts: a general context (demographic aspects and economic and welfare aspects) and an educational context (investment in Education, the organization of education, and above all, the impact of learning[2]).

Curriculum design and development

As a core element of any education system, curriculum design and development is too substantial to be passed over for evaluation. However, it is a complex task for evaluation experts.

[(2)] We propose that the impact of learning be measured using the Educational Capabilities Index developed by Bonal, Rambla, Calderón and Pros (2005) and updated by Álvarez – López for the 2013/2014 academic year. (Álvarez – López, 2017a).

Received: 29-05-2018 Accepted: 12-04-2019

At the Autonomous Community level, the SEM evaluation of curriculum design focuses on four main aspects: educational objectives, contents, competencies, and learning standards. In contrast, curriculum development should be evaluated fundamentally through methodologies, measures for attending to diversity, educational resources and, of course, teacher coordination between subjects and departments. While it is admittedly not easy to evaluate methodologies at the systemic level, the technical challenge should not be used as a hindrance to improving the quality of the system.

Teachers

One of the most urgent aspects to evaluate is the teachers. While it is true that it generates resistance, it would be of great help to teachers if it is done for purposes of diagnosis and improvement and has no negative consequences, and if the teachers themselves, their representatives, and professionals of recognized prestige take part in its design. A teacher evaluation of this kind would contribute to the improvement of the system.

It is proposed that this evaluation address lifelong learning (courses, incentives, applicability of any training received, training organizations, modalities and teacher satisfaction) and teacher performance. It is essential that teacher performance be approached from several sides. For example, it should evaluate the teaching methodology in relation to the subject area of the courses he or she teaches and how well he or she adapts that method to the harmonious development of the competencies that students are supposed to acquire. In addition, teacher performance includes the work of tutoring, accompanying and advising students and their families. Another element to assess is his or her identity as a teacher, answering questions such as "How do you perceive your work in relation to society?" or "What do you think society expects of teachers?"

An internal evaluation of the students themselves, of their families, of the fellow teachers and of the management team must be incorporated into the teacher evaluation. Only a complete view of their teaching work can correctly dimension their contributions to the education system.

Received: 29-05-2018 Accepted: 12-04-2019

Side effects

The very existence of education systems has a set of side effects on the society in which they are immersed. There are at least three very relevant side effects of education systems that must be taken into account[3]: 1) on the job market (direct or indirect creation of jobs); 2) on the local economy (due to the numerous additional services generated by education: transportation, meal service, supplies, materials, maintenance, etc. and the taxes derived from these activities); and 3) on social inclusion and integration.

Taken together, these side effects lead to an increase in the social productivity of education systems in addition to the educational productivity that they create.

School management teams

School management is a key factor to quality teaching as well as to improving school organization and performance. Evaluation of the management function is therefore a necessary condition for improving the work of the management teams and, in turn, the organization and functioning of the schools.

Accordingly, a program for evaluating the management function should at least address the management and coordination of the school's activity, the strengthening of the school's collaboration with families, the municipality and other institutions, the promotion of a school climate that favors study, the engagement of participation, the promotion of peaceful coexistence at the school, the development of innovation programs, and the participation of the faculty in the curriculum.

Students

Evaluations of students from an external point of view are much more limited than evaluations carried out from inside the classroom.

(3) To develop this dimension, we relied on the interviews with Guillermo Gil carried out in the framework of a research stay in the INEE.

Received: 29-05-2018 Accepted: 12-04-2019

Consequently, what is needed is a more thorough evaluation of achievement, taking into account a competency paradigm in which all competencies are equally important, and all must be addressed if students are to be evaluated correctly. However, student performance is only a product; processes are also important. Therefore, the evaluation must also measure the student's commitment to the task and to studying as well as the learning style. Finally, it would be useful to evaluate student participation in school affairs in school-led initiatives and their degree of satisfaction with the school life.

Inspection

As a fundamental pillar in the proper functioning of the education system, inspections should not be overlooked in the model. Evaluating the Inspectorate of Education, however, is not easy to accomplish. In this case, the proposed model checks the degree of compliance with the assigned functions, and the degree of satisfaction of the schools, management teams, teachers and guidance teams with the work of the Inspectorate. It would be highly worthwhile for the educational community to assess the work of the Inspectorate and its usefulness.

Moreover, a number of different aspects of the Public Administration may be analyzed, ranging from evaluating the regional and provincial structures of the Inspectorate down to the selection processes of inspectors.

Policies

Just as there are authors who understand evaluation of the education system as a type of program evaluation, it may be assumed that an evaluation of the education system is already a de facto evaluation of education policies. However, from the systemic perspective proposed here, education policies will be treated as elements in their own right.

The academic literature is extensive in terms of public policy evaluation (e.g., Fisher, Miller and Sidney, 2006; Nagel, 2001). The topics of evaluation include the quality of public services, cost-benefit analysis,

analysis of implementation, risk assessment, and evaluation of the impact of a policy, etc[4].

Programs and projects

The evaluation of programs and projects is a reality in many regions in Spain. This section includes all the initiatives of an educational nature that arise from the public administration, external to the schools and designed by the administration and executed by schools willing to participate in them. Any project implemented at a regional level must be evaluated in terms of its design, implementation, impact and user satisfaction.

Programs and projects addressing diversity also deserve a closer look. Because of their importance in guaranteeing the effective right to education for all people, it should be compulsory for education to design follow-up evaluations of such programs.

Resources

The purpose of this type of evaluation is to analyze how resources should be used to optimize results, encourage quality teaching and promote pedagogical innovation.

The most classical way to evaluate resources is by analyzing investment in education through two types of indicators: relative investment (usually estimated as the percentage of GDP invested in education) and absolute investment (investment in terms of real monetary value). These investments, relative and absolute, are mainly translated into concrete investments in personnel, material, and financial resources for the schools, as well as in other activities to support and promote education.

Resource evaluation includes assessing the effectiveness of resource allocation and management by taking into account expenditures, source of resources, distribution mechanisms, management capacity and

(4) There are many approaches we could take in this regard, but without a doubt any public policies on Education should include the work of the former Agencia Española de Evaluación y Calidad (Spanish Evaluation and Quality Agency), part of the Ministry of Internal Revenue and Public Function and created in 2007 under the name of State Agency for the Evaluation of Public Policies and the Quality of Services.

Received: 29-05-2018 Accepted: 12-04-2019

monitoring mechanisms, mainly covering the areas of human resources, leadership, back-office staff, physical and material resources, and specific educational programs.

Relationship between the school and families and society

It is important to know how schools relate to their surroundings and families. In this sense, the school-society relationship should be approached by how well the school is integrated into its surroundings and how much society participates in school life, and vice-versa. On the other hand, the school-family relationship is a more delicate and important issue. The participation of families in the functioning of the schools is a crucial to the success of the teaching-learning processes. Elements that can be analyzed in this regard include which channels of participation exist aside from the ones established by regulations (e.g., the school board and the PTA), and how participation in these channels is articulated and how satisfied the parties involved are with their usefulness.

The justification for incorporating these processes into the evaluation of the educational system is settled by the need to coordinate with the families, since good relations and even complicity among educational agents must be encouraged for the student and child to develop in harmony. In this sense, Bolívar (2006), taking up Bourdie's proposal, advocates building up social capital in the educational community.

Results

Talk about results most often leads to a discussion of student achievement. However, the outcomes of education cover a much broader area that may best be served by addressing the achievement of the purposes of the education system. Accordingly, it has become increasingly common to evaluate educational achievement, especially its external evaluation, as an evaluation suited to all purposes of education and applicable to all levels and fields. This, however, is far from being methodologically rigorous. For example, regarding Article 2 of the Consolidated LOE, a large-scale written test cannot possibly evaluate "education in the

Received: 29-05-2018 Accepted: 12-04-2019

exercise of tolerance and freedom within the democratic principles of coexistence, as well as in the prevention of conflicts and their peaceful resolution" (Article 2c of the Consolidated LOE).

Therefore, evaluations are needed that can analyze the degree of fulfillment of the purposes and objectives of the education system and the effectiveness and efficiency of the system beyond the academic achievements of the students.

Support services and psycho-pedagogical intervention

Although support services may vary between Autonomous Communities, all should be taken into account when assessing the education system. Some services are external to the schools, such as psycho-pedagogical teams (in primary school), teacher training centers, pedagogical consultancies (for example, the Basque *berritzegunes*), the Technical Specialists of Community Services, etc. Other services are internal to the schools, such as the Guidance Department at high schools, the services of Hearing and Language or Speech Therapy, Therapeutic Pedagogy or even Physiotherapy, and other workers such as the Technical Educational Assistant (at least in Castilla-León). All these workers, and any others that can be found in the different regional realities, must be incorporated insofar as the quality of the provision of their services as well as the satisfaction of the people involved. Designing evaluation programs on these services, knowing their situation, diagnosing possible areas for improvement and implementing processes of change, restructuring, incorporation of resources, etc., are measures that may introduce greater levels of equality in the system, which will certainly contribute significantly to improving its quality.

Category 3. Gathering and analyzing information

Indicators

Indicator systems are a useful tool for policymakers, a way to get an overview of the system synthetically. To make this tool useful in a systemic

evaluation, the indicators should be selected taking into account context, inputs, processes and results. It is important for there to be a balance between indicators, without over-dimensioning any of them.

However, the SEM proposal discourages explanations of the data that arise from the selected indicators, because the data can only be interpreted correctly if they are accompanied and complemented with other elements. Therefore, a system of indicators - however complete it may be - should never attempt to replace the evaluation programs, which will always be more complete.

Instruments and strategies for collecting information

There is a sort of unanimity regarding the need to have both instruments -for the more quantitative conceptions of data analysis- and strategies -for the most qualitative conceptions of data analysis- because they provide different information (Bisquerra, 2004). In practice, however, instruments (such as questionnaires and tests) have always taken precedence over strategies (such as interviews, discussion groups or life histories). The problem has resided in there being limited resources to address strategies throughout the breadth of the education system.

The SEM model proposed here advocates effective combination of both types of analysis. They must coexist and complement each other, as the information they provide is not the same, and they can help explain each other. Thus, the incorporation of qualitative data collection strategies entails the need to train evaluators in both information collection strategies and analysis techniques.

The advisability of having strategies makes it necessary to come up with solutions to the cost of resources. In this sense, the evaluations should be more sample-based rather than census-based. Census evaluations at the systemic level are difficult to justify if they prevent collecting qualitative information from the system. They may make sense in the early years of an evaluation system to collect information from all the parties involved and thus be able to plot out trends with less margin for error. However, once the trends have been identified, it may suffice for subsequent evaluations to follow up using samples. Therefore, the census nature may be sacrificed in favor of incorporating information-gathering strategies, taking on a greater margin of random error.

The same can be said of the annual nature of some evaluations, but this will be dealt with under the heading of timeliness.

Category 4. Dissemination and use of results

Dissemination of results

Depending on how the program is designed, dissemination will take place through specific channels and to specific audiences. However, the results from any evaluation of the education system must then be made available to three different groups: politicians, who will disseminate them to society, schools, which will work on them with teachers, and families, who are the main stakeholders.

It is commonly accepted that the most appropriate channel of information is the results report. This report adapts the information and language to fit the target audience. Furthermore, it must be manageable in its size, structure and presentation of the results. The Regional Evaluation Units (REUs) should take pains to design simple but rigorous reports.

Moreover, the report must be purely informative. The explanation of results (the explanatory nature) and the proposal of measures to improve results (the propositional nature) should not be the REU's job; it should first and foremost be the job of the parties involved, and in the last case, of politicians and society. On some very specific occasion an interpretation of the results may be justified, but always with the caution of being partial.

In the SEM, the report must account for the results from all the partial evaluations of the different dimensions, while also maintaining its systemic nature.

Along with the report, it is important to hold sessions to present the results. These meetings should be held with any collectives of the system that are directly or indirectly affected by the evaluation -at their express request- and in any case informative sessions should be held with the media, in the regional parliament, and with interested members of society (in open session).

In addition, the public report must be available in the information channels that the Educational Administration has, such as their web page or the information subscription networks that they manage. There they

Received: 29-05-2018 Accepted: 12-04-2019

should remain so as to facilitate the longitudinal reading of the results and promote their analysis by researchers.

Use of results

The SEM is designed in the understanding that there can be no evaluation without decision-making. The results from evaluations of the educational system should be discussed and analyzed with the educational community through the relevant representative bodies and with those who have regional government representation.

Once all the representatives of the education system have been consulted in different forums, and taking into account the reflections and proposals that may have emerged, any final decision-making will have to be shaped by the people in charge of each partial evaluation program of the system as well as the political and technical leaders of the education administration so that decision-making is made as participatory as possible. This way ensures two things: on the one hand, that the entire educational community be informed of the state of the educational system, be able to propose solutions to problems and feel like an active part of the process. On the other hand, it ensures that the decisions made have the consensus needed for smooth and effective implementation. Although it is a slower process, the reward is social peace and calm changes.

In no case do we consider rankings to be of any interest in evaluating the educational system. The only desirable comparison in education is with oneself, with respect to the previous moment. This applies to the student, the teacher, the school, the Autonomous Community and the State itself.

Category 5. Time

This section refers to two aspects: a) the moment for collecting data and for publishing the results and b) the frequency with which evaluations are carried out.

Regarding the first aspect, it is generally agreed that there is no perfect moment. This would depend on the meaning and purpose of

the evaluation. In any case, if the evaluation is census-based, the results should be published with enough time to include improvements in the following school year. Sample-based evaluations do not force this immediacy because there is no individual analysis or results of centers or students.

In the SEM it is unlikely that the complete evaluation process (Figure III) can be carried out in a single school year. Student evaluations have the additional handicap of it being preferable to do them at the end of the course so that the assessed competence is as developed as possible. Fortunately, this problem is easily solved by ensuring evaluations do not take place on an annual basis; if not, perhaps biannually. It is preferable for schools to apply improvements that are not directly linked to the students and teachers of a particular school year (thereby taking some of the pressure off them) rather than continually receiving external evaluations. Furthermore, schools already carry out their evaluations (some even have teacher and school self-evaluations), discuss the results of educational evaluations, and seek improvements for the problems they detect without the need for the Administration to intervene. Consequently, the information that an external evaluation can offer them is something they often already know. This information is most useful for the evaluation of the education system.

On the other hand, if evaluation programs are carried out for the different dimensions every two or three years, their joint analysis should be carried out over a longer period of time, perhaps three to five years. This would give enough time for all the partial evaluations to be incorporated, and the results and proposals for improvement to be discussed unhurriedly. As a result, a complete evaluation of the educational system would be available every three to five years, a very manageable period in the times in which Education is developed.

In short, it is important to space the evaluations because the system does not change from one year to the next and can easily become overloaded. It is a waste of resources and energy that may be necessary in other sections of the systemic evaluation. Therefore, we propose that evaluations be biannual or triennial, and that the overall evaluation of the education system be a three- to five-year process, with the participation of all involved.

Synthesis of the Systemic Evaluation Model

The complete diagram of the model features the elements that come into play and the relationships between them, but fails to show the importance that processes have in the model. Therefore, to make it easier to understand, below is a schematic summary of the phases of the SEM (Figure III).

The figure shows the evaluation process for any dimension of the system, from the design to the suggestions for improvement. In addition, it can help understand the process of the complete evaluation of the education system if, as of Phase 4, it includes all the reports from the partial evaluations and all the people (or their representatives) involved in the analysis, decision-making, and suggestions for improvement.

FIGURE III. Phases of the SEM implementation process

Source: Prepared by the authors

To conclude: Discussion and recommendations for the proper functioning of the SEM

Discussion of the model

After the presentation of the SEM, it is time to propose some points for discussion and to point out limitations and future lines of research.

In the first place, the SEM provides new considerations with respect to earlier models. By providing considerable comprehensiveness on the whole system, this new model goes beyond the functionality, effectiveness, and efficiency of the system that De la Orden proposes as fundamental elements of evaluation. Moreover, in relation to Gargallo's proposal, preference is given to the professional development of teachers, a conceptualization that introduces aspects such as their performance. All this gives the model sufficient entity to be implemented, bearing in mind that its confrontation with each specific reality where it is implemented will be a determining issue. The same may be said of student evaluations; earlier models are updated here by incorporating the competency paradigm (which includes both results and processes). Another contribution of the SEM model as opposed to these authors' models is that it is articulated within the framework of each regional government context.

Secondly, the SEM is not without its limitations. It is a theoretical model that, as has been pointed out, must be contrasted with reality. While its design takes into account the contributions of the political and technical specialists of the Autonomous Communities that participated in the study, it has not yet been contrasted with the educational reality, with school times, with regional legislation or with the functioning of the regional evaluation units. Furthermore, the SEM is the result of a study that had its own methodological limitations (Álvarez - López, 2017a) that, if overcome, could help improve its design.

Finally, the spirit of the SEM reveals a trend towards a greater presence of evaluation in the education system. In this regard, the SEM runs the risk of legitimizing evaluations as instruments for merely monitoring the system rather than for improving it. If the definition of systemic evaluation proposed here is applied correctly and the following recommendations are taken into account, this risk could be brought under control.

Recommendations for the SEM to function correctly

The recommendations are presented below in two groups. In one are the recommendations referring to educational administration, all the aspects that should be taken into account at the level of public administration to accompany the Systemic Evaluation Model. In the other are those

that refer to policies and actions aimed at supporting and promoting the model and endowing it with resources[5].

Educational Administration

The main recommendations for educational administration are the following:

- Approval of a regional educational law that grants legislative security to the policies for evaluating the educational system.
- Autonomy of the evaluating institution: REUs on the agency level and high degrees of political - institutional and technical - methodological autonomy.
- Coordination between educational administration and REUs: from the Ministry of Education, Culture and Sports as well as from the regional Departments of Education and the Inspectorate. In addition, it would be advisable to generate channels of information and debate between REUs to prevent the "mirror wall" effect (Álvarez - López, 2017b).
- Organizationally flexible and versatile REUs. For this flexibility and versatility to make full sense, there must be a clear ongoing training plan for workers (in aspects such as methodologies or the dissemination of information).
- REU managers with technical expertise and direct links with research centers and universities
- Formation of a Scientific Committee in the REUs as an advisory body with variable composition
- List of positions with steady jobs, no service commissions, to reduce temporary employment in the educational system and give stability to the REU.
- Encouraging educational community to participate in evaluations, from design to decision-making with the results.
- Generating synergies from the REU with schools and school management teams.

(5) For a more comprehensive discussion, see Álvarez – López (2017a)

Policies

The recommendations for policies, initiatives, and specific actions are intended as suggestions for improving the functioning of the SEM. Although there may not be a direct cause - effect relationship, or the impact may not be immediate, they all have the success of the model and thus the improvement of the quality of education as their ultimate goal.

The main policy recommendations are as follows:

- Predictable, sustained funding of both the REUs themselves and the evaluation programs.
- Directly transferred financial resources for the REUs that foster the autonomy of the evaluating entities in the expenditure.
- Ministerial projects with finalist financing: much the same as the PROA Plan to encourage the evaluation of elements that are not currently being evaluated.
- Promoting general evaluation plans to generate certainty and security in the education system.
- Creating pedagogical consultancy units much like the Basque *berritzegunes*, which can be especially important when designing improvement plans.
- Promoting an evaluative culture that generates trust in the evaluators and normality in the evaluative act. Evaluation meant to improve the system, not to audit it.

Future steps in the framework of this proposal would be to contrast the SEM in areas of regional reality and to begin gathering evidence that can help validate it.

Bibliographic References

Álvarez – López, G. (2017a). *La evaluación de la educación básica en el sistema educativo español: estudio comparado en el ámbito de las Comunidades Autónomas*. Tesis Doctoral. Universidad Autónoma de Madrid.

Received: 29-05-2018 Accepted: 12-04-2019

Álvarez – López, G. (2017b). Descentralización educativa y evaluación de sistemas educativos: conceptualización y análisis del caso español. *RIESED, International Journal of Studies in Educational Systems*, 7 (2), pp. 52 – 73.

Bertalanffy, L. von (1969). *General System Theory: Foundations, Development, Applications* (Revised Edition). New York: George Braziller Inc.

Bisquerra, R. (coord.) (2004). *Metodología de la investigación educativa*. Madrid: La Muralla.

Bolívar, A. (2006). Familia y escuela: dos mundos llamados a trabajar en común. *Revista de Educación*, 339, pp. 119 – 146.

Bolívar, A. (2016). ¿Cómo puede la evaluación institucional contribuir para mejorar la escuela? *Estudos em Avaliaçao Educacional*, *27* (65) pp. 284 – 313.

Bonal, X., Rambla, X., Calderon, E. y Pros, N. (2005). La descentralización educativa en España. Una mirada comparativa a los sistemas escolares de las Comunidades Autónomas. Estudis 18. Barcelona: Fundació Carles Pi i Sunyer.

Caballero, Á., Manso, J., Matarranz, M. y Valle, J. M. (2016). Investigación en Educación Comparada: pistas para investigadores noveles. *Revista Latinoamericana de Educación Comparada*, 39, 39 – 56.

Calero, J. y Choi, A. (2012). La evaluación como instrumento de política educativa. *Presupuesto y Gasto Público*, 67, 29 – 42

Colom, A. J. (1979). *Sociología de la educación y teoría general de sistemas*. Barcelona: Oikos – Tau Ediciones.

Davies, P. (2004). *Is evidence – based government possible? 4th Annual Campbell Collaboration Colloquium*. Washington D.C. Retrieved from: http://www.sandycampbell.com/sc/KTC_Module_1_files/KTC%20Module%201%20-%202.3b%20%20Davies%202004.pdf

Eurydice (2015). *Assuring Quality in Education. Policies and Approaches to School Evaluation in Europe.* Luxembourg: Publications Office of the European Union.

Fisher, F., Miller, G. J. & Sidney, M. S. (Eds.) (2006). Handbook of Public Policy Analysis: Theory, Politics, and Methods (Public Administration and Public Policy). Florida: Routledge.

García Garrido, J. L. (1996). *Fundamentos de Educación Comparada.* (3ª ed.). Madrid: Dykinson.

Received: 29-05-2018 Accepted: 12-04-2019

Gargallo, B. (2003). Un modelo de evaluación del sistema educativo. *Revista Española De Pedagogía,* 61 (224), pp. 33 – 59

Hanushek, E. A., y Raymond, M. E. (2001). The confusing world of educational accountability. *National Tax Journal, 54*(2), 365-384.

Maestro, C. (2006). La evaluación del sistema educativo. *Revista de Educación*, nº extraordinario, 315 – 336

Nagel, S. (2001). *Handbook Public Policy Evaluation*. California: SAGE.

Orden, A. de la (1997). Desarrollo y validación de un modelo de calidad universitaria como base para su evaluación. *Revista Electrónica de Investigación y Evaluación Educativa*, *3*(1–2). Retrieved from: https://www.uv.es/RELIEVE/v3n1/RELIEVEv3n1_2.htm

Orden, A. de la (2009). Evaluación y calidad: Análisis de un modelo. *ESE. Estudios sobre Educación*, 16, pp. 17 – 36.

Orden, A. de la y Jornet, J. M. (2012). La utilidad de las evaluaciones de sistemas educativos: El valor de la consideración del contexto. *Bordón. Revista de Pedagogía, 64*(2), pp. 69-88.

Reynolds, D., Chapman, C., Kelly, A., Muijs, D. y Sammons, P. (2012). Educational effectiveness: The development of the discipline, the critiques, the defence, and the present debate. *Effective Education, 3*(2), 109-127. doi:10.1080/19415532.2011.686168

Tiana, A. (2014). Veinte años de políticas de evaluación general del Sistema educativo en España. *Revista de Evaluación de Programas y Políticas Públicas*, 2, 1 – 21

Valle, J.M. (2008). Valores y pluralismo social en la formación para la convivencia ciudadana. In Sociedad Española de Pedagogía, Educación, ciudadanía y convivencia. Diversidad y sentido social de la educación. In Touriñan, J.M. (Dir.) *Actas del XIV Congreso Nacional y III Iberoamericano de Pedagogía* (pp. 129-149). Madrid: SEP.

Valle, J. M. (2013). Supranational Education: a new field of knowledge to address educational policies in a global world. *Journal of Supranational Policies of Education*, 1, 7 – 30.

Contact address: Gabriel Álvarez-López. Universidad Autónoma de Madrid, Facultad de Profesorado y Educación, Departamento de Didáctica y Teoría de la Educación. Avda. / Francisco Tomás y Valiente, 3. 28049, Madrid. E-mail: g.alvarez@uam.es

Received: 29-05-2018 Accepted: 12-04-2019

Attitudes toward Statistics in Education Students: Profile Analysis

Actitudes hacia la Estadística en Alumnos de Educación: Análisis de Perfiles

DOI: 10.4438/1988-592X-RE-2019-385-421

Xavier G. Ordóñez Camacho
Universidad Complutense de Madrid
Sonia Janeth Romero Martínez
Universidad a distancia de Madrid
Covadonga Ruiz de Miguel
Universidad Complutense de Madrid

Abstract

Attitudes towards statistics are a series of behaviours, cognitions and emotions revolving around this field. Attitudes are learned, mediated by cultural factors, which influence students' learning and performance. Thus, the purpose of the present study is to describe these attitudes in a sample of 855 university students from the Faculty of Education at the Complutense University of Madrid, by identifying profiles that also include variables on identification, past education, academic self-concept and their assessments of statistics in their careers and professional futures. The methodology used combines factor analysis and cluster analysis methods. The procedure included the application of two questionnaires (one to measure attitudes and the other for all other variables), the identification of the number of factorial components and groups and profile validation and description. Four groups were identified: the first two (61.64% of the sample) are comprised primarily of women in Preschool / Primary Education who revealed unfavourable attitudes, high anxiety, low self-concept and who consider statistics of little use professionally. The other two groups are made up of students with favourable attitudes, high self-concepts and who believe that statistics are quite useful. The present study represents a first step in identifying and characterising

Received: 23-07-2018 Accepted: 12-04-2019

profiles on attitudes towards statistics among Education students, in order to create Education degrees to arouse interest in the subject among the groups that most need it.

Keywords: attitudes, statistics, cluster analysis, profiles, university students

Resumen

Las actitudes hacia la estadística son un cúmulo de conductas, cogniciones y emociones respecto a la estadística. Las actitudes son aprendidas, mediadas por factores culturales e influyen en el aprendizaje y rendimiento de los alumnos, por ello, el objetivo del presente trabajo es realizar una descripción de dichas actitudes en una muestra de 855 alumnos universitarios de la Facultad de Educación de la Universidad Complutense de Madrid por medio de la identificación de perfiles que incluyen también variables de identificación, formación previa, auto-concepto académico y valoraciones sobre la estadística en su carrera y futuro profesional. La metodología que se ha utilizado combina métodos factoriales y análisis cluster. El procedimiento incluye la aplicación de dos cuestionarios (uno para medir las actitudes y otro para las demás variables), la identificación del número de componentes factoriales y de grupos y la validación y descripción de los perfiles. Se han identificado cuatro grupos: los dos primeros (61.64% de la muestra) están compuestos principalmente por mujeres del grado Educación Infantil/Primaria que presentan actitudes desfavorables, alta ansiedad, bajo auto-concepto y consideran la estadística poco útil profesionalmente. Los otros dos grupos están compuestos por alumnos que presentan actitudes favorables, alto auto-concepto y consideran que la estadística es bastante útil. El presente estudio constituye un primer paso en la identificación y caracterización de perfiles de actitud hacia la estadística en alumnos de Educación y permite crear programas educativos para despertar el interés hacia la materia en los grupos que más lo necesitan.

Palabras clave: actitudes, estadística, análisis de conglomerados, perfiles, alumnos universitarios.

Introduction

Studying attitudes towards statistics has been a fundamental topic in Spain in recent years, especially after the Bologna Process was implemented, making this subject compulsory in curricula due to its importance when systematising, analysing and interpreting data.

It is common to find students in the classroom who think they cannot handle this subject due to its difficulty. This makes them pay less attention to contents, commit less to their studies and have a superficial relationship with the subject (León and Vaiman, 2013). Some studies show that students approach the subject with negative preconceptions and attitudes (Aparicio and Bazán, 2005; Carmona, 2004), and have stated that the way in which students handle it is fundamental to their success or failure (Evans, 2007; Vanhoof et al., 2006).

That is why the objective of this paper is to understand and discover the attitudes with which students face studying statistics, with the aim of identifying different profiles to thus help to improve teaching practices. By identifying profiles, the aim is to achieve a complete description, not only of attitudes toward statistics, but also with regard to variables on identification, previous education, academic self-concept and the inclusion of statistics in their professional futures and careers.

The study of attitudes toward statistics is relatively new, starting in the 1980s (Blanco, 2008; Estrada, 2009; Gal, Ginsburg and Shau, 1997; Gómez, 2010; Mondéjar, Vargas, and Bayot, 2008; Mondéjar and Vargas, 2010). The attitude toward statistics is a trend that forms over time and as a consequence of the emotions and feelings experienced in a learning context. It could be defined as a sum of emotions and feelings, stable and resistant to change, that slowly develop and in which cultural factors play an important role (Gal et al., 1997). Rodríguez Feijoo (2011) defines it as a learned and enduring organisation of beliefs and cognitions, endowed with an emotional load and predisposing one to action coherent with these cognitions and emotions.

Three components are commonly identified in its composition: cognitive, emotional and behavioural (Darías-Morales, 2000), where the emotional load is the main weight, followed by cognitive and, finally, behavioural with a lesser weight (Estrada, 2009).

Previous research on the topic has centred on the construction and adaptation of psychometric tools to measure the construct of attitude as a predictor of performance in the subject and on variables associated with attitude development. Theoretical models explaining attitude and its effects on variables like students' interests and goals have been handled to a lesser degree (Blanco, 2011; Bourne and Nesbit, 2018).

With regard to attitude as a predictor of performance, there are several studies that have proven a positive relationship between attitude and

performance in statistics (Evans, 2007; Rosli, Maat and Rosli, 2017; Sesé et al., 2015). The factors identified as predictors of attitudes are grouped into three categories: past experience in the area, anxiety and self-beliefs surrounding performance (self-concept, self-efficacy, expectations of success), where past experience tends to have the greatest predictive weight on later performance. Ashaari, Judi, Mohamed and Tengku Wook Wook (2011) and Auzmendi (1992) also identify factors like expectations of success-failure at the beginning, attitude towards technology, previous objective education, motivation and anxiety.

With regard to the relationship between attitudes and other variables, these results can be found: with regard to gender, there are contradictory results, some authors finding that women have a higher level of anxiety and more negative attitudes towards mathematics (Cuesta, Rifá and Herrero, 2001), while others do not find differences (Baloglu, 2003). The age variable also shows contradictory results: some studies find a relationship between these variables (Katz and Tomazic, 1988) and others conclude that older students have higher anxiety levels about statistics (Baloglu, 2003).

With regard to individuals' self-perception of their capacities, this last variable has been found to condition attitudes and anxiety (Ashaari, Judi, Mohamed and Tengku Wook, 2011; Kottke, 2000). Other authors (Benson 1989; Bandalos and Yates, 1995) find that the higher the self-concept is for mathematics, the lower the anxiety level is for statistics. More recent studies have also discovered this relationship (Jaiswal and Choudhur, 2017; Obilor, 2012).

Works that study the attitudes of students with different degrees were also found: Tourism (Cladera et al., 2019), Medicine (Mat-Din et al., 2018); Psychology (Chiesi and Primi, 2017), Teaching (Pulido, 2009), Physical Activity and Sport Sciences (Tejero-González and Castro, 2011), languages (Cimpoeru and Roma, 2018) and also among secondary school students (Ashaari, Judi, Mohamed and Tengku Wook, 2011; Huynh, 2018).

For prior studies on attitudes done using cluster analysis, we found the paper by Cárdenas (2008), which identifies types of attitudes towards mathematics among secondary school students, finding four groups of students. Judi, Ashaari, Mohamed and Tengku Wook (2011) perform discriminant analyses with university students. This study finds a majority of students (70% of the sample) who have favourable attitudes towards statistics. And, more recently, Saki, Tabesh, Yousefi and Khalili (2017) did

Received: 23-07-2018 Accepted: 12-04-2019

a cluster analysis in Iran, finding three distinct groups of students, the first with a good attitude and high competence, the second with a poor attitude and low competence, and the third with low competence but a high interest in learning.

Method

Sample

A total of 855 university students enrolled in Applied Statistics or Research Methods in Education classes from four different degree programmes participated: Pedagogy (129), Preschool Education (299), Primary Education (317) and Social Education (110) at the Complutense University of Madrid. Of this number, 85.8% of the individuals were taking the course for the first time. The sample taken was not random, as the instrument was applied to all students in these subjects during the 2013-2014 academic year. Students participated voluntarily and, via informed consents, they were assured that the results would be handled ethically and anonymously.

The sample was 84% women and 16% men. The age range of study participants varies from 17 to 65 years old, with a mean age of 20.47 (SD=3.65).

Variables

The variables in this research are summarised in Table I.

TABLE I. Variables defined for the present study

Type	Abbreviation	Variable	Response Options
Identification	P1	Degree	Pedagogy Preschool Education Primary Education Social Education
	P2	Age	Continuous
	P3	Gender	Male Female
	P4	First time taking the statistics class	Yes No
Previous Education	P5	Coming from	Baccalaureate (BACH) Vocational Training (VT)
	P6	Option studied	Science and Technology (SC and TECH) Humanities and Social Sciences (HUM and SS) Arts
	P7	Assessment of knowledge of statistics before starting degree programme	Zero Basic Intermediate Advanced Expert
	P8	Knowledge of presence of statistics in the degree curricula	Yes No
	P9	Have taken a subject with a math component (Maths, Statistics) during the last year of baccalaureate or vocational module	Yes No
	P10	If so, mark obtained	Takes values between 0 and 10
	P11	Difficulty of studying subjects with a math component in the past	None Little Normal Quite a bit A lot

Received: 23-07-2018 Accepted: 12-04-2019

Self-perception	P12	What capacity do you think you have for subjects related to numbers?	None Little Normal Quite a bit A lot
	P13	What current knowledge do you have of statistics?	Zero Basic Intermediate Advanced Expert
	P14	Do you think you will pass the Statistics class here?	Yes, in February Yes, but in September I don't think I will pass it
	P15	What do you think your mark (numerical) will be in the statistics class?	Takes values between 0 and 10
	P16	What relationship do you think a statistics class has with maths?	None Little Quite a bit A lot
	P17	What level of anxiety does the statistics class cause you?	Takes values between 1 and 10
	P18	How interesting is the statistics class to you?	Takes values between 1 and 10
	P19	How useful is the statistics class to you?	Takes values between 1 and 10
Professional future	P20	What use do you think statistics will have for you in your career?	None Little Quite a bit A lot
	P21	When you finish your degree, would you like to work in a field related to research, in which you would have to handle/apply statistics?	Yes No I haven't thought about it
	P22	When you finish your degree, would you like to specialise in Statistics, doing for example of Master's of Research?	Yes No I haven't thought about it
Statistics in the degree	P23	Statistics is a subject that should be removed from this degree programme	Likert 5 point scale (strongly disagree – strongly agree)
	P24	There should be more statistics training in this degree programme	
	P25	If I had the chance, I would enrol in more statistics classes	
	P26	Using computer programs makes it easier to understand the subject	
	P27	Using real data helps to understand the subject better	
Attitudes towards statistics (QATS)	P28	Total score	Continuous
	P29	NE score	
	P30	PE score	
	P31	U score	

Source: author elaboration

Received: 23-07-2018 Accepted: 12-04-2019

Of the 31 variables defined for the study, variables P2, P10, P15, P17, P18, P19, P28, P29, P30 and P31 are taken as continuous variables and the rest as categorical variables.

Instruments

A questionnaire was used to gather all the variables in Table 1. Attitudes toward statistics were measured via the Questionnaire of Attitudes Toward Statistics (QATS; Ordoñez, Romero and Ruiz de Miguel, 2016), comprised of 16 items rated on a 5 point scale from strongly disagree (1) to strongly agree (5).

The QATS was used to estimate the subjects' scores in three areas: Negative Emotions (NE) made up of 8 questions, Positive Emotions (PE) comprised of 4 questions, and Usefulness (U) also consisting of 4 questions. The instrument shows proof of construct validity and high reliability (Ordoñez, Romero and Ruiz de Miguel, 2016).

Data analysis

Analysis was conducted with an analysis technique that combines factor analysis and cluster analysis methods (Lebart, Morineau and Piron, 1995) in four analysis phases: In the first a Multiple Correspondence Analysis (MCA) was performed that let us transform categorical variables into continuous ones (Husson, Lê and Pagès, 2010; Lebart, Morineau and Piron, 1995; Pardo and Del Campo, 2007).

During the second phase, we did hierarchical clustering using the Ward method (Pardo and Del Campo, 2007). The third phase was classification via mobile centres using the K-means method. The groups identified were also validated by hypothesis testing.

The last phase was to describe each of the groups using continuous and categorical variables. If the variables are continuous, the mean of each group is compared to the overall average and if categorical, the percentage of each group is compared to the overall percentage. To determine if they are statistically significant, the value test statistic was employed (Husson, Josse and Pagès, 2010).

Received: 23-07-2018 Accepted: 12-04-2019

Data analysis was performed with R, version 3.2.4 (R Development Core Team, 2016) and for MCA and Cluster Analysis the FactoMineR package, was employed (Lê, Josse and Husson, 2008) version 1.32 (Husson, Josse, Le and Mazet, 2016).

Results

Identification of the number of components

First, a MCA was done with the 16 items on the QATS, of which three components were kept, as that is the number of scales making up the QATS. Selection criteria for the components were proposed for the factor analysis (Dray, 2008; Josse and Husson, 2012; Saccenti and Camacho, 2015), but not for the MCA, for which the selection criterion continued to be theoretical assessment of the components by the researcher. Table II summarises the results of the eigenvalues for the components.

TABLE II. Components selected and their eigenvalues according to MCA

Component	Eigenvalues	Percentage of variance	Cumulative percentage of variance
1	0.44	10.99	10.99
2	0.28	6.89	17.88
3	0.16	4.09	21.97

Source: author elaboration

Identification of the number of groups

In the dendrogram presented in Figure I one can see that the 855 students can be classified into four groups: the first is comprised of 165 (19.30%), the second 362 (42.34%), the third by 214 (25.03%) and the last by 114 (13.33%) students.

Received: 23-07-2018 Accepted: 12-04-2019

FIGURE I. Dendrogram depicting the 4 groups identified

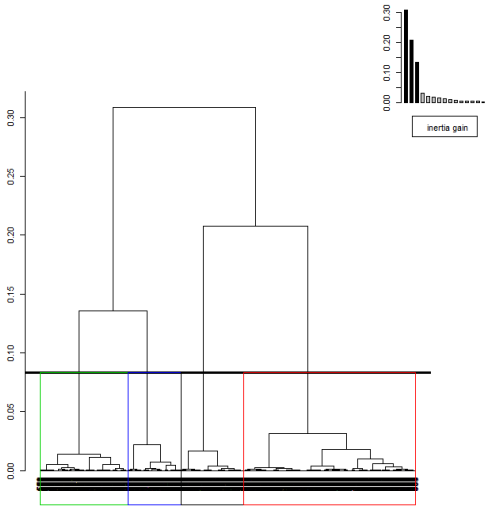
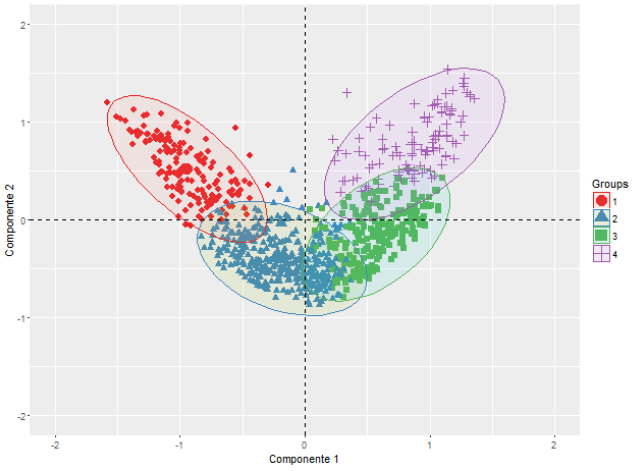


Figure II shows the individuals and the group to which they belong plotted on the first factorial plane.

FIGURE II. Individuals plotted on the first factorial plane



Received: 23-07-2018 Accepted: 12-04-2019

Validation of the identified groups

Hypothesis testing on the differences between the four groups identified by the variables defined in Table I let us confirm that identification is valid, since according to the Kruskal-Wallis test, there are differences in all quantitative variables (Table III). The effect size (r) of these differences was estimated with the Mann-Whitney's U test, employing the Rosenthal formula (1991). Low sizes are observed in age and previous mark values ($r \leq 0.30$), while high ones are seen in scores on the QATS scales ($r \geq 0.70$). With regard to categorical variables, a chi-squared test was done, where the results show significant differences between groups, except for variables P3 and P16 (see Table IV).

TABLE III. Kruskall-Wallis and effect sizes for continuous variables

Variables	Mean				SD				Levene's Test[a]		x^2 Kruskal-Wallis[b]	Effect size (r)					
	1	2	3	4	1	2	3	4	F	p		1-2	1-3	1-4	2-3	2-4	3-4
Age	21.15	20.22	20.10	20.99	2.96	3.72	3.39	4.45	1.82	0.14	23.73	0.19	0.22				
Mark obtained	6.83	6.83	7.15	7.21	0.80	0.92	0.98	1.05	2.42	0.07	42.74		0.24	0.30	0.18	0.21	
Mark expected	5.24	5.77	6.51	6.92	1.04	0.90	0.91	0.86	0.99	0.40	239.66	0.27	0.57	0.68	0.38	0.49	0.23
Anxiety	8.70	6.81	4.96	3.50	1.47	1.82	2.04	1.98	9.78	0.00	390.05	0.50	0.74	0.81	0.43	0.57	0.33
Interest	4.01	5.35	6.22	6.57	2.16	1.88	1.64	1.96	6.51	0.00	140.00	0.29	0.50	0.53	0.25	0.28	
utility valuation	3.72	5.13	5.93	5.84	2.05	1.92	1.82	2.09	4.50	0.00	108.69	0.30	0.48	0.44	0.21	0.15	
NE score	35.71	27.23	18.54	12.36	3.47	3.55	3.04	2.70	6.69	0.00	710.49	0.72	0.86	0.85	0.78	0.74	0.72
PE score	4.92	6.30	8.95	10.07	1.74	2.02	2.46	3.58	41.45	0.00	324.78	0.38	0.72	0.70	0.50	0.46	
U score	11.27	13.56	15.28	15.57	3.89	2.90	2.53	2.66	11.81	0.00	157.48	0.29	0.52	0.54	0.30	0.29	
Total score	28.48	40.63	53.69	61.28	4.85	5.05	4.55	6.10	5.20	0.00	708.88	0.73	0.86	0.85	0.80	0.74	0.55

a. df_1=3 y df_2 = 851

b. df=3 y p<0.0001

Source: author elaboration

TABLE IV. Chi-squared for categorical variables

Variables	x^2	*df*	*p*
EAE2[a]	801.05	12	0.00
EAE4[a]	329.82	12	0.00
EAE9[a]	192.06	12	0.00
EAE12[a]	884.90	12	0.00
EAE14[a]	404.47	12	0.00
EAE17[a]	1236.31	12	0.00
EAE18[a]	814.39	12	0.00
EAE19[a]	239.21	12	0.00
SATS2[a]	1169.59	12	0.00
SATS5[a]	180.08	12	0.00
SATS6[a]	642.79	12	0.00
SATS7[a]	101.03	12	0.00
SATS10[a]	149.67	12	0.00
SATS11[a]	1024.78	12	0.00
SATS19[a]	132.75	12	0.00
SATS21[a]	925.13	12	0.00
P1	38.64	9	0.00
P3	10.97	6	0.09
P4	13.69	6	0.03
P5	17.65	9	0.04
P6	37.28	9	0.00
P7	113.01	9	0.00
P8	17.16	3	0.00
P9	36.83	3	0.00
P11	253.62	15	0.00
P12	261.52	12	0.00
P13	136.07	9	0.00
P14	95.28	6	0.00
P16	8.74	12	0.72
P20	129.25	12	0.00
P21	124.98	6	0.00
P22	83.50	6	0.00
P23	308.10	12	0.00
P24	169.05	12	0.00
P25	168.83	12	0.00
P26	44.39	12	0.00
P27	58.20	12	0.00

QATS test items

Source: author elaboration

Received: 23-07-2018 Accepted: 12-04-2019

Description of the first student profile (group 1)

Table V shows the categorical variables and Table VI the continuous variables that characterise the students identified with the first profile.

TABLE V. Categorical variables that describe group one

Variable (Response Options)	Cla/Mod	Mod/Cla	Global	v.test[a]
P23(Agree)	60.98	45.45	14.39	11.29
P24(strongly disagree)	39.18	63.64	31.35	9.59
P11(A lot)	50.00	41.21	15.91	8.99
P12(Little)	37.17	60.61	31.46	8.66
P25(strongly disagree)	28.35	87.27	59.42	8.58
P21(No)	28.54	84.24	56.96	8.23
P20(None)	57.33	26.06	8.77	7.70
P7(Zero)	33.73	50.91	29.12	6.60
P12(None)	70.97	13.33	3.63	6.33
P13(Zero)	52.31	20.61	7.60	6.22
P22(No)	23.03	93.94	78.71	5.84
P14(Yes, but in September)	43.62	24.85	10.99	5.76
P14(I don't think I will pass it)	62.96	10.30	3.16	5.01
P9(No)	26.21	65.45	48.19	4.95
P27(strongly disagree)	72.22	7.88	2.11	4.84
P13(Basic)	24.88	63.64	49.36	4.09
P1(Preschool Education)	26.42	47.88	34.97	3.80
P5(VT)	25.96	44.85	33.33	3.43
P8(No)	23.12	69.09	57.66	3.33
P26(disagree)	37.04	12.12	6.32	3.14
P25(strongly agree)	71.43	3.03	0.82	2.87
P24(strongly agree)	50.00	3.64	1.40	2.35
P11(Quite a bit)	23.66	37.58	30.64	2.12
P3(Woman)	20.71	84.85	79.06	2.07
P23(4)	25.34	22.42	17.08	1.98

a. $p<0.05$

Source: author elaboration

In Table V, one can see that the group is made up of women, in the Preschool Education degree and who came from VT, who state that they have had great difficulty with mathematics studies in the past, that

they have not taken a course with a maths component during their last academic year and that they did not know that the curricula included the subject of statistics. Further, they ranked their previous and present knowledge of statistics as zero, and also believe they will not pass the class in the first exam sitting, that they are not interested in jobs at which they would have to apply statistics, believe that statistics are not useful for their professional performance and are not interested in specialising in statistics. They also believe that statistics should be removed from the degree syllabus, are not considering enrolling for any other statistics courses and are not interested in using real data or computer programs for learning.

TABLE VI. Continuous variables that describe group one

Variable	Mean in category	Overall mean	Sd in category	Overall sd	v.test[a]
NE score	35.71	24.71	3.46	8.17	19.23
Anxiety	8.70	6.27	1.46	2.47	14.06
Age	21.15	20.47	2.95	3.63	2.66

a. $p<0.05$
Source: author elaboration

According to the results of Table VI, the students in the first profile are characterised by having a very high mean on NE with regard to the overall mean. The same is true of Anxiety and Age.

Description of the second student profile (group 2)

Table VII depicts the categorical variables that describe the group and Table 8 the continuous variables.

Received: 23-07-2018 Accepted: 12-04-2019

TABLE VII. Categorical variables that describe group two

Variable (Response Options)	Cla/Mod	Mod/Cla	Global	v.test[a]
P11(Quite a bit)	56.11	40.61	30.64	5.39
P23(3)	57.39	36.46	26.90	5.37
P22(No)	45.77	85.08	78.71	3.94
P23(4)	56.85	22.93	17.08	3.86
P21(No)	47.23	63.54	56.96	3.33
P24(2)	49.39	44.48	38.13	3.26
P25(strongly disagree)	46.85	65.75	59.42	3.23
P20(Little)	47.02	61.05	54.97	3.06
P12(Normal)	47.68	53.87	47.84	3.02
P7(Basic)	46.39	53.31	48.65	2.33
P13(Basic)	46.21	53.87	49.36	2.26
P26(3)	48.35	32.32	28.30	2.22
P1(Primary Education)	46.69	40.88	37.08	1.97

a. $p<0.05$
Source: author elaboration

In Table VII we see that this group is characterised by studying Primary Education, stating they have had quite a bit of difficulty in the study of maths courses in the past, assess their prior knowledge of statistics as basic, their current knowledge also as basic and believe that their capacities for subjects related to numbers are normal. Likewise, they are not interested in jobs at which statistics would have to be applied, believe that statistics has little use for their career development and are not interested in specialising in statistics. They agree that statistics should be removed from the degree programme and disagree that there should be more statistics training in the degree. They would not enrol in other statistics courses and are neutral about the use of computer programs that would help them understand the subject.

TABLE VIII. Continuous variables that describe the second clusters

Variable	Mean in category	Overall mean	Sd in category	Overall sd	v.test[a]
NE score	27.23	24.71	3.54	8.17	7.73
Anxiety	6.81	6.27	1.82	2.47	5.49

a. $p<0.05$
Source: author elaboration

According to the results shown in Table VIII, the students in group two are characterised by having a high NE with regard to the total average, although it is lower than the mean for group one (see Table III). With regard to anxiety, the mean is also lower than for group one (see Table III).

Description of the third student profile (group 3)

Table IX depicts the categorical variables that describe the group and Table X the continuous variables.

TABLE IX. Categorical variables that describe group three

Variable (Response Options)	Cla/Mod	Mod/Cla	Global	v.test[a]
P25(2)	42.20	42.99	25.50	6.54
P7(Intermediate)	45.73	35.05	19.18	6.48
P24(2)	43.46	38.79	22.34	6.40
P11(Normal)	37.37	50.47	33.80	5.84
P22(I haven't thought about it)	43.11	33.64	19.53	5.75
P21(I haven't thought about it)	36.00	54.67	38.01	5.72
P24(3)	39.90	38.79	24.33	5.50
P13(Intermediate)	35.01	55.14	39.42	5.38
P14(Yes, in February)	27.93	95.79	85.85	5.28

Received: 23-07-2018 Accepted: 12-04-2019

P20(Quite a bit)	35.14	45.33	32.28	4.62
P12(Quite a bit)	41.27	24.30	14.74	4.35
P11(Little)	38.21	21.96	14.39	3.50
P27(4)	32.68	38.79	29.71	3.30
P25(3)	38.32	19.16	12.51	3.25
P9(Yes)	29.57	61.21	51.81	3.18
P12(Normal)	29.58	56.54	47.84	2.94
P23(strongly disagree)	33.94	26.17	19.30	2.86
P8(Yes)	29.56	50.00	42.34	2.60
P3(Man)	34.11	20.56	15.09	2.51
P21(Yes)	41.86	8.41	5.03	2.47
P1(Pedagogy)	33.33	20.09	15.09	2.30
P13(Advanced)	41.94	6.07	3.63	2.08
P26(4)	29.75	33.64	28.30	1.98

a. $p<0.05$
Source: author elaboration

As seen in Table IX, this group is primarily made up of men in the Pedagogy degree. They claim they have had little difficulty with the study of maths subjects in the past and had knowledge that the curriculum included the subject. They rated their prior knowledge as intermediate, their current knowledge as advanced and believe their capacities for related subjects are high. They also think they will pass the course on the first exam sitting, are interested in jobs at which they would have to apply statistics, consider it useful for their careers and have not thought about whether they would be interested in specialising in statistics yet. They do not agree that statistics should be removed from the degree programme and agree with using real data and computer programs to understand the subject better.

TABLE X. Continuous variables that describe group three

Variable	Mean in category	Overall mean	Sd in category	Overall sd	v.test[a]
Total score	53.69	44.31	4.54	11.83	13.39
PE score	8.95	7.20	2.45	2.95	10.01
Mark expected	6.51	6.01	0.91	1.08	7.85
U score	15.28	13.82	2.52	3.35	7.36
Utility valuation	5.93	5.15	1.82	2.09	6.28
Interest	6.22	5.47	1.63	2.07	6.10
Mark obtained	7.15	6.96	0.98	0.94	3.34

a. $p<0.05$

Source: author elaboration

Table X shows the means of the total QATS score, NE and U scores, where the assessment of Usefulness, Expected Mark, Previous Mark and Interest statistically differ from the mean of the entire sample, which means these students are characterised by having more favourable attitudes toward statistics, as well as having higher marks obtained in the past and a higher expected mark than the total sample.

Description of the fourth student profile (group 4)

Table XI presents the categorical variables that describe the group and Table XII the continuous variables

Received: 23-07-2018 Accepted: 12-04-2019

TABLE XI. Categorical variables that describe group four

Variable (Response Options)	Cla/Mod	Mod/Cla	Global	v.test[a]
P23(strongly disagree)	31.52	45.61	19.30	6.96
P11(Little)	34.15	36.84	14.39	6.53
P12(Quite a bit)	30.95	34.21	14.74	5.67
P25(3)	30.84	28.95	12.51	5.11
P14(Yes, in February)	15.12	97.37	85.85	4.32
P24(3)	22.60	41.23	24.33	4.29
P26(strongly agree)	21.79	44.74	27.37	4.28
P20(Quite a bit)	20.65	50.00	32.28	4.21
P6(Science and Technology)	23.90	33.33	18.60	4.06
P21(I haven't thought about it)	18.77	53.51	38.01	3.60
P13(Advanced)	38.71	10.53	3.63	3.59
P12(A lot)	45.00	7.89	2.34	3.47
P21(Yes)	32.56	12.28	5.03	3.33
P9(Yes)	16.93	65.79	51.81	3.22
P1(Social Education)	23.64	22.81	12.87	3.17
P22(Yes)	46.67	6.14	1.75	3.12
P27(strongly agree)	16.78	64.04	50.88	3.02
P7(Advanced)	34.62	7.89	3.04	2.79
P22 (I haven't thought about it)	19.76	28.95	19.53	2.61
P13(Intermediate)	16.32	48.25	39.42	2.05

a. $p<0.05$
Source: author elaboration

In Table XI we can see that this group is characterised by studying Social Education and coming from baccalaureate studies. They state that they have had little difficulty with the study of maths subjects in the past and believe that their prior knowledge of statistics is advanced, current knowledge is also advanced and believe they have high capacities for the subject. They also think they will pass the course in the first exam sitting and are interested in a job at which they would have to apply statistics, believe that statistics will be quite useful in their careers and are interested in specialising in statistics. They do not think the subject should be

removed from the degree curriculum and they think that real data and computer programs should be used to better understand the subject.

TABLE XII. Continuous variables that describe group four

Variable	Mean in category	Overall mean	Sd in category	Overall sd	v.test[a]
Total score	61.28	44.31	6.07	11.83	16.45
PE score	10.07	7.20	3.56	2.95	11.15
Mark expected	6.92	6.01	0.85	1.08	9.69
Interest	6.57	5.47	1.95	2.07	6.10
U score	15.57	13.82	2.65	3.35	6.00
utility valuation	5.84	5.15	2.08	2.09	3.78
Mark obtained	7.21	6.96	1.05	0.94	3.03

$p<0.05$

Source: author elaboration

According to the results shown in Table XII, the variable means of the QATS total score, scores for NE, U, Usefulness Assessment, Expected Mark, Previous Mark and Interest differ significantly from the total mean of all individuals. Further, the means of these variables are higher than those seen in group three (see Table X). Thus, the students in this group are characterised by having very favourable attitudes to statistics.

Discussion and conclusions

As a whole, the results prove three important conclusions: on the one hand, the extraction of the groups suggests that there are two student profiles with unfavourable attitudes toward statistics and its educational usefulness, where the groups show several combinations of variables, although they share a low academic self-concept for statistics, do not think they will use statistics in their future jobs and have negative perceptions about statistics in their careers, where the differential features occur in the identification variables. Students in group one are older

Received: 23-07-2018 Accepted: 12-04-2019

and it has also been longer since they studied subjects related to maths, their anxiety level is also very high and they have negative emotions related to statistics. This result is in line with that which was found by Carmona (2004) and Gal, Ginsburg and Schau (1997), who highlight the importance of past experience, anxiety and academic self-concept in the development of favourable attitudes toward statistics. The present study also confirms that, as group one, with more negative attitudes, is also the group with the worst past experience, more negative self-beliefs and a high level of anxiety about the subject.

Secondly, although groups one and two contain the majority of the sample (61.64%), two student profiles were found with positive attitudes, although they also differ in some aspects. With regard to common characteristics, these students have studied related subjects, come from a technological baccalaureate and have a good academic self-concept, rating their knowledge of statistics as advanced. They are students who value the inclusion of statistics in their professional futures and have positive perceptions of statistics in their careers. One issue that merits mention is that the students in these two profiles agree with using real data and computer programs to better understand the subject, thus confirming the relationship between a positive attitude and a less superficial relationship with the subject (León and Vaiman, 2013).

Third, the present study represents an initial step in the identification and characterisation of student profiles, including individual variables, past education, academic self-concept and perceptions on the incorporation of statistics in their careers and professional futures. Authors including Bandalos and Yates (1995) and Benson (1989) have analysed the relationships between some of these variables, finding that the higher the self-concept, the more favourable the attitudes are toward statistics, although these authors do not consider variables such as using statistics in professional futures and during their careers, making the contributions of the present study considerable.

Previous and expected marks appear in the characterisation of groups three and four, as this variable does not appear in groups one and two, indicating that the latter two do not have positive expectations of their performance in the subject. The present study does not include the real mark obtained in the subject, because our interest was not in predicting performance, but instead to explore and describe the profiles. However, it would be worth adding this variable with a predictive nature in future studies.

With regard to the characterisation of the groups as far as age, gender and degree, we found that the first two groups, characterised by having a negative attitude, were composed primarily of women, while the third group, with a positive attitude, was comprised of men. This result matches that which was found by Cuesta, Rifá and Herrero (2001), who discovered a higher level of anxiety and more negative attitudes toward mathematics among women.

With regard to age, this study reveals that students in the first profile have an average age higher than the overall sample, which coincides with studies by Baloglu (2003) and Katz and Tomazic (1988), who also found that older students have higher levels of anxiety toward statistics. And with respect to the degree, although works were found that studied the attitudes of students in different degree programmes like Psychology (Chiesi and Primi, 2017), Teaching (Pulido, 2009) and Physical and Sport Sciences (Tejero-González and Castro, 2011), a specific study was not found within the specific areas of Education with which to compare the findings herein. However, it is interesting that the first two profiles are made up of students in Preschool-Primary Education, while the other two groups are made up of students in Pedagogy and Social Education, a result that could be researched in future studies.

The validity of the groups was verified statistically, but also confirmed by their commonality with groups found in the studies by Cárdenas (2008). Students in groups T1 and T2 in Cárdenas's study (2008) have a low self-concept, high anxiety and a lack of understanding of mathematics language and are, therefore, quite similar to groups one and two of the present study. However, the T2 in the study by Cárdenas (2008) is also characterised by thinking that mathematics is important for their future, whereas we did not find mixed profiles that combined high and low levels of variables. Students in the T4 group showed a neutral profile that has not been found in the present research either. One possible explanation of this result (which could be checked in future research projects) is that as time goes by, attitudes become more radical, as Cárdenas's study (2008) was done with students in basic education, while this research was done with university students.

Along the same line, Judi, Ashaari, Mohamed and Tengku Wook (2011) identified (in a sample of university students) two clearly different groups with positive and negative attitudes. The results of the aforesaid study are very similar to those obtained here, and validate the groups

presented here, as we also found that the group with favourable attitudes is characterised by enjoying the course, believing in their high capacities and recognising the benefits of statistics in the professional life, while the group with unfavourable attitudes is characterised by their lack of interest.

With regard to methodological issues, when we analyse the relationships between the groups and the variables, one can notice that the cluster analysis represents a tool that goes beyond simple relationships with a linear association, letting the profiles be fully characterised and, thus, generate initiatives in the future focused on the groups that most need them.

Although the results of the present study represent an advance in scientific knowledge, cluster analysis does have an inherent limitation, as its nature depends on the sample being analysed, which is why the results of the present research cannot be generalised to the population as a whole. In order for the profiles found here to be applicable to the general population, future studies will need to be conducted that find similar groups to thus establish firm conclusions on the relationships between all variables. Although important similarities were found with studies by Cárdenas (2008) and Judi, Ashaari, Mohamed and Tengku Wook (2011), we must proceed with caution as the samples used are different.

Despite the aforesaid limitation, the present study does represent a first step in identifying and characterising profiles on attitudes toward statistics among Education students, letting innovative and well-founded education degree programmes be created to spark interest in the subject among the student groups that most need it.

In conclusion, we can confirm that the QATS lets subjects at risk of failing the *statistics* subject be identified in order to implement different support strategies with these individuals, such as student-teacher tutoring (physical and virtual), heterogeneous working groups (the identification of profiles permits this) and looking for support at university guidance services, via the study on attitude-based variables (anxiety, motivation, etc.) and other factors (study habits, work planning, self-regulation, etc.) related to the study.

References

Ashaari, N. S., Judi, H. M., Mohamed, H. y Tengku Wook, T. M. (2011). Student's Attitude towards Statistics Course. *Procedia Social and Behavioral Sciences, 18*, 287-294. DOI:https://doi.org/10.1016/j.sbspro.2011.05.041

Auzmendi, E. (1992). *Las actitudes hacia la matemática- estadística en las enseñanzas medias y universitarias.* Bilbao: Mensajero.

Baloglu, M. (2003). Individual differences in statistics anxiety among college student. *Personality and Individual Differences*, *34*, 855-865. DOI:https://doi.org/10.1016/S0191-8869(02)00076-4

Bandalos, D. y Yates, K.C. (1995). Effects of math self-concept, perceived self efficacy and attributions for failure and success on test anxiety. *Journal of Educational Psychology, 87*(4), 611-623. DOI:10.1037/0022-0663.87.4.611

Bazán, J. y Aparicio, A. (2006). Las actitudes hacia la matemática-estadística dentro de un modelo de aprendizaje. *Revista de Educación de la Pontificia Universidad Católica del Perú*, *15*(28), 7-20.

Benson, J. (1989). Structural components of statistical test anxiety in adults: An exploratory model. *Journal of Experimental Education*, *57*, 247-261.DOI:http://dx.doi.org/10.1080/00220973.1989.10806509

Blanco A. (2008). Una revisión crítica de la investigación sobre las actitudes de los estudiantes universitarios hacia la Estadística. *Revista Complutense de Educación, 19*(2), 311-330. DOI:http://dx.doi.org/10.5209/RCED

Blanco A. (2011). Applying social cognitive career theory to predict interests and choice goals in statistics among Spanish psychology students. *Journal of Vocational Behavior, 78,* 49–58. DOI: 10.1016/j.jvb.2010.07.003

Bourne y Nesbit (2018). Do Attitudes towards Statistics Influence the Decision to Study Psychology at Degree Level? *Psychology Teaching Review, 24*(2), 55-63.

Cladera, M., Vich-i-Martorell, G., Rejón, F. y Juaned, C. (2019). Tourism students' Attitudes Toward Statistics. *Journal of Hospitality, Leisure, Sport y Tourism Education, 24*, 202–210. DOI:https://doi.org/10.1016/j.jhlste.2019.03.002

Cimpoero, S. y Roman, M. (2018). Statistical Literacy and Attitudes Towards Statistics of Romanian Undergraduate Students. MPRA.No.90452. Disponible online:https://mpra.ub.uni-muenchen.de/90452/

Cárdenas, C. S. (2008). Identificación de tipologías de actitud hacia las matemáticas en estudiantes de séptimo y octavo grados de educación secundaria. *Perfiles educativos,30*(122), 94-108.

Carmona, J. (2004). Una revisión de las evidencias de fiabilidad y validez de los cuestionarios de actitudes y ansiedad hacia la estadística. *Statistics Education Research Journal*, *3*(1), 5-28.

Chiesi, F. y Primi, C. (2017). Do attitudes toward statistics change during an introductory statistics course? A study on Italian psychology students. CERME10 (Feb. 2017). Dublin, Ireland.

Cuesta, M., Rifá, H. y Herrero, F. (2001). *Un estudio exploratorio, en estudiantes de psicología, de una escala de actitudes hacia la estadística*. VII Congreso de Metodología de las Ciencias Sociales y de la Salud, Madrid, 25-28 septiembre.

Darías-Morales, E. (2000). Escala de actitudes hacia la Estadística. *Psicothema, 28*, 175-178.

Dray, S. (2008). On the number of principal components: A test of dimensionality based on measurements of similarity between matrices. *Computational Statistics y Data Analysis*, *52* (4), 2228-2237. DOI: 10.1016/j.csda.2007.07.015

Estrada, A. (2009). *Las actitudes hacia la estadística en la formación de los profesores*. Lérida: Milenio.

Evans, B. (2007). Student attitudes, conceptions and achievement in introductory undergraduate college statistics. *The Mathematics Educator*, *17*(2), 24-30.

Gal, I.; Ginsburg, L. y Schau, C. (1997). Monitoring attitudes in learning statistics. towards an Assesment Framework. En I. Gal, y J. Garfield (eds.). *The Assesment Challenge in Statistics Education*. Amsterdam, Holanda: IOS Press and International Statistical Institute.

Gómez, E. (2010). Actitudes de alumnos de postgrado hacia la estadística aplicada a la investigación. *Revista Encuentro XLII*, *85*, 27-38.

Husson F.; Lê S. y Pagès J. (2010). *Exploratory Multivariate Analysis by Example. Using R.* Londres, Inglaterra: Chapman y Hall.

Husson, F.; Josse, J.; Le, S. y Mazet, J. (2016). *FactoMineR: Factor analysis and data mining with R. R package version 1.32.* Paris, Francia. http://CRAN.R-project.org/package=FactoMineR

Huynh, M. (2018). *Bringing the islands to schools: improving student attitudes towards statistics.*Tesis doctoral. RMIT University. Australia.

Josse, J. y Husson, F. (2012). Selecting the number of components in principal component analysis using cross-validation approximations. *Computational Statistics y Data Analysis, 56* (6), 1869-1879. DOI:10.1016/j.csda.2011.11.012.

Jaiswal, K. y Choudhuri, R. (2017). Academic Self Concept and Academic Achievement of Secondary School Students. American Journal of Educational Research, 5(10),1108–1113.DOI:10.12691/education-5-10-13.

Judi, H. M.; Ashaari, N. S.; Mohamed, H. y Tengku Wook, T. M. (2011). Students profile based on attitude towards statistics. *Procedia Social and Behavioral Sciences, 18*, 266-272. DOI:https://doi.org/10.1016/j.sbspro.2011.05.038

Katz, B. M. y Tomazic, T. Z. (1988). Changing student's attitudes toward statistics through a nonquantitative approach. *Psichological Reports*, *62*, 658. DOI:http://dx.doi.org/10.2466/pr0.1988.62.2.658

Kottke, J. (2000). Mathematical proficiency, statistics knowledge, attitudes toward statistics, and measurement course performance. *College Student Journal*, *34*(3), 334-347.

Lê, S.; Josse, J. y Husson, F. (2008). FactoMineR: an R package for multivariate analysis. *Journal of statistical software*, *25*(1), 1-18.DOI: 10.18637/jss.v025.i01.

Lebart, L.; Morineau, A. y Piron, M. (1995). *Statisquique exploratoire multimensionnelle*. París: Dunod.

León, E. y Vaiman, M. (2013). Actitudes, experiencia previa y nivel de logro en Estadística en la carrera de Psicología. En G. Contreras, M. Cañadas y P. Arteaga (eds.). *Actas de las Jornadas Virtuales en Didáctica de la Estadística, Probabilidad y Combinatoria*. Granada, España.

Mat-Din, H., Arifin, W.N., Abdullah, S., Yusoff, M.S.B., Abd-Aziz, A. The internal structure of new attitudes towards statistics inventory among postgraduate students in medical and health sciences. Education in Medicine Journal. 2018;10(4):53–67. DOI:https://doi.org/10.21315/eimj2018.10.4.6

Mondéjar, J. y Vargas, M. (2010), Determinant factors of attitude towards quantitative subjects: differences between sexes. *Teaching and teacher education*, *26*(3), 688-693. DOI:10.1016/j.tate.2009.10.004.

Mondéjar, J.; Vargas, M. y Bayot, A. (2008). Medición de la actitud hacia la estadística. Influencia en los procesos de estudio. *Revista electrónica de investigación psicoeducativa*, *16* (3), 729-748.

Obilor, I. (2012). Relationship Between Self-Concept and Mathematics Achievement of Senior Secondary Students in Port Harcourt. *Journal Plus Education, 8*(1), 169-178. Disponible en: //www.uav.ro/jour/index.php/jpe/article/view/928/988

Ordoñez, X.; Romero, S. y Ruiz de Miguel, C. (2016). Cuestionario de actitudes hacia la estadística (CAHE): evidencias de validez y fiabilidad de las puntuaciones en una muestra de alumnos de educación. *Bordón: Revista de Pedagogía*, *68*(4), 121-136. DOI: 10.13042/Bordon.2016.41826.

Pardo C. E. y Del Campo P. C. (2007). Combinación de métodos factoriales y de análisis de conglomerados en R: el paquete FactoClass. *Revista Colombiana de Estadística*, *30*, 235–245.

Pulido, J. E. (2009). Enseñanza de la estadística a partir de la actitud del alumno. *Laurus. Revista de Educación*, *15*(30), 42-70.

R Development Core Team (2016). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org

Rodríguez Feijoo, N. (2011). Actitudes de los estudiantes hacia la estadística. *Interdisciplinaria*, *28*(2), 199-205.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* Newbury Park (CA): Sage

Rosli, M. K., Maat, S. M., y Rosli, R. (2017). Students' attitude and anxiety towards statistics: A descriptive analysis. *Research on Education and Psychology, 1*(1), 47-56.

Saccenti, E. y Camacho, J. (2015). Determining the number of components in principal components analysis: A comparison of statistical, crossvalidation and approximated methods. *Chemometrics and Intelligent Laboratory Systems*, *149*, 99-116. DOI: https://doi.org/10.1016/j.chemolab.2015.10.006G

Saki, A., Tabesh, H., Yousefi, R. y Khalili, S. (2017). Clustering the Attitudes towards Statistics and Technology among Medical Post Graduate Students. *British* Journal *of Applied Science & Technology, 18*(4), 1-11. DOI:10.9734/BJAST/2016/29673

Sesé, A., Jiménez, R., Montaño, J.J. y Palmer, A. (2015). Can Attitudes toward Statistics and Statistics Anxiety Explain Students' Performance?

Revista de Psicodidáctica, 20(2), 285-304.DOI: 10.1387/RevPsicodidact.13080

Tejero-González, C. y Castro-Morera, M. (2011). Validación de la escala de actitudes hacia la estadística en estudiantes españoles de Ciencias de la actividad física y el deporte. *Revista Colombiana de Estadística*, *34*(1), 1-14.

Vanhoof, S.; Castro, A.; Onghena, P.; Verschaffel, L. y Van Dooren, W. (2006). Attitudes toward statistics and their relation with short and long term exam results. *Journal of Statistics Education, 14*(3), 1-5. DOI:https://doi.org/10.1080/10691898.2006.11910588

Contact address: Xavier Giovanni Ordóñez Camacho, Universidad Complutense de Madrid, Facultad de Educación, Departamento de Investigación y Psicología en Educación. Edificio La Almudena. Rector Royo Villanova s/n. Ciudad Universitaria. 28040, Madrid. E-mail: xavor@ucm.es

Received: 23-07-2018 Accepted: 12-04-2019

Is death in the Spanish curriculum?[1]

¿Está la muerte en el currículo español?

DOI: 10.4438/1988-592X-RE-2019-385-422

Agustín de la Herrán Gascón
Pablo Rodríguez Herrero
Victoria de Miguel Yubero
Universidad Autónoma de Madrid

Abstract

This article presents a diagnostic research study on the inclusion of death awareness in the Spanish curriculum. Its documentary analysis methodology examines legislation on the Spanish education system and its curricular development at the state level as contexts of analysis. Analysis of the presence of death awareness in the curriculum is based on 25 units: terms and lexical roots selected by 17 experts. Quantitative techniques were used to analyse their frequency in the current laws on education and the curriculum at different education levels (Pre-school, Primary, Compulsory Secondary and High School), areas of knowledge and subjects, where appropriate. The qualitative analysis interprets the presence of each term, according to the inclusion death awareness in education. The results allow for identifying the presence of curricular elements that are comparable to the Pedagogy of Death in all education levels and in areas of knowledge such as Natural Sciences, Social Sciences, Humanities and Arts, though not in a planned manner, that is, with the intention of educating in death awareness. Likewise, its absence is observed in the aims and principles of Education or in key competences. The map resulting of the study allows for concluding the existence of educational gaps and challenges that must be addressed by including death awareness as an essential element of education.

Key words: curriculum, pedagogy, awareness, death, finitude, documentary research

---

(1) "Funded by R+D+i call 'Retos de la Sociedad' 2017, reference: EDU2017-85296-R (AEI/FEDER, UE)."

Received: 04-11-2018 Accepted: 12-04-2019

Resumen

Este artículo presenta una investigación de diagnóstico respecto a la inclusión de la conciencia de muerte en el currículo español. Se realiza a través de una metodología de tipo documental, considerando como contextos de análisis la legislación referente al sistema educativo español y su desarrollo curricular a nivel estatal. Las unidades de análisis, desde las que se analiza la presencia de la conciencia de muerte en el currículo, son 25 términos y raíces léxicas, seleccionados por 17 expertos. El examen de los términos en la normativa se realiza a través de técnicas cuantitativas y cualitativas. Desde las primeras se analiza la frecuencia de su presencia en las leyes educativas vigentes y el currículo de las distintas etapas educativas (infantil, primaria, secundaria obligatoria y bachillerato), áreas de conocimiento y materias, en su caso. El análisis cualitativo interpreta la presencia de cada término, de acuerdo al planteamiento de la inclusión de la conciencia de muerte en la educación. Los resultados permiten detectar la presencia de elementos curriculares asimilables a la Pedagogía de la muerte en todas las etapas educativas y en áreas de conocimiento como Ciencias de la Naturaleza, Ciencias Sociales, Humanidades y Artes, si bien no de manera planificada, es decir, con la intención de educar en la conciencia de muerte. Asimismo, se observa su ausencia en los fines y principios de la educación o en las competencias clave. El mapa que arroja la investigación deduce lagunas educativas y retos de futuro asociados a una educación que incluya la conciencia de muerte como elemento indispensable para la formación.

Palabras clave: currículo, pedagogía, conciencia, muerte, finitud, estudio documental

Introduction

The purpose of this study is to determine the presence of death awareness in education-related regulations and in the Spanish curriculum of Pre-school, Primary, Compulsory Secondary and High School on a state level. The goal of this documentary study is to empirically determine whether or not national laws and curriculum address death awareness. If it does, then it is more likely that the issue will be included in official education and in classrooms.

Initially, we are concerned with death as an educational construct for two reasons: because it is part of life and, therefore, it is not possible to educate for life without including death (Herrán and Cortina, 2006).

Received: 04-11-2018 Accepted: 12-04-2019

Furthermore: "Education and learning centres must address, foremost, what is most important to humankind" (Rodríguez, Herrán and Cortina, 2015a, p. 191), and death is significant, for obvious reasons, to any conscious being.

Many perspectives may justify the presence of death in education: its profound interest for everyone, its educability throughout the lifetime, its impact on self-awareness, maturity, sensibility, cultural socialisation, improved coexistence, preparedness for significant losses, etc. The essential question, according to the researchers, is how education is possible without death awareness. As Verdú (2002) affirms, teachings that omit death disqualify any institution of knowledge.

The baseline is the construct of the Pedagogy of death, understood as an emerging discipline whose object of study is education, teaching and training that include the death awareness. This includes two educational approaches (Cortina and Herrán, 2011; Herrán and Cortina, 2006; Herrán, González, Navarro, Freire and Bravo, 2000): the curricular approach that seeks to normalise death in its teachings and learnings, and the palliative approach by which tutors offer an educational accompaniment during situations of mourning.

Research in the field of the Pedagogy of death dates back to the decade of the 20s of the last century (Rodríguez, Herrán and Cortina, 2019), when researchers from Anglo-Saxon countries (Pine, 1977) began studying lines of training in healthcare settings through the field they referred to as *Death Education*. Its irruption in education dates to the 80s, in Spain (Fullat, 1982; Mèlich, 1989) and other countries (i.e., Berg, 1982; Fertziger, 1983; Sahler, 1983) alike.

Since then, we can differentiate distinct lines of research on the Pedagogy of death: (a) Approaches and dimensions of the Pedagogy of death, and its association with the life cycle and education on emotions (Aspinall, 1996), the students' comprehensive development (Corr, Nabe and Corr, 2000), the sociocritical perspective (Mantegazza, 2004), the radical and inclusive focus based on death awareness (Herrán and Cortina, 2006; Petitfils, 2016; Rodríguez, Herrán and Cortina, 2015b) or the pedagogy of genocide, war and the Holocaust (Bos, 2014; Lindquist, 2007; Mariana, 2016; Zembylas, 2011); (b) Effects of the Pedagogy of death, through studies that have analysed the impact of education programmes for death, in relation to the decrease of fear (Berg, 1982), anxiety (Glass, 1990), eradication of violence in schools (Jaramillo, 2017;

Shackelford, 2003) or education on values (Hulbert, 1999); (c) Didactics of death, which studies the training-related value of films (Cortina and Herrán, 2011), partial deaths and small deaths (Herrán et ál., 2000), the service-learning methodology (Rodríguez et ál., 2015a), stories (Colomo, 2016) or music (Colomo and Oña, 2014); (c) Accompaniment in situations of mourning at school, highlighting the design and evaluation of educational activities for implementation in situations of mourning (i.e., Kroen, 2002; Rodríguez et ál., 2015b; Turner, 2004) or the analysis of the tutor as a figure (Herrán et ál., 2000; Herrán and Cortina, 2008).

To date, we have not found any empirical research that approaches the presence of death awareness in the curriculum, despite references on the importance of including death in the education system curriculum (Herrán, González, Navarro, Freire and Bravo, 1998; Jackson and Colwell, 2002), including pre-school (Herrán et ál., 2000; Herrán and Cortina, 2006), primary (Edgar and Howard-Hamilton, 1994; Herrán and Cortina, 2006) and secondary (Hetzel, Winn and Tolstoshev, 1991; Herrán and Cortina, 2006) education levels.

Given its brief lifetime on both international (Rodríguez et ál., 2019) and national (Rodríguez, Herrán and Cortina, 2012) levels, the Pedagogy of death may be considered, strictly speaking, at the frontier of our knowledge on education. One mandatory point of departure is to explore it in the Spanish curriculum and in applicable laws on education, as this type of study has neither been implemented in Spain yet nor, in the manner proposed, in another education system.

Method

Research design

This study has been implemented using a documentary methodological design. The method was chosen for its value in analysing official and primary documents on the phenomenon object of the study (Payne and Payne, 2004). In our case, it will study the presence of death in the Spanish curriculum on a state level. An analysis is performed of official regulations and Pre-school, Primary, Compulsory Secondary and High

Received: 04-11-2018 Accepted: 12-04-2019

School education. These documents define the basic curriculum statewide in the education levels that the study encompasses.

As of the research design, a combined analysis complements the interpretation of both quantitative and qualitative data. The quantitative analysis includes a descriptive approach to the frequency and percentage of the presence of the selected terms, given their relationship with the Pedagogy of death. The qualitative analysis focuses on the contents by coding and interpreting the contextual significance of the terms associated with the presence of death included in the curriculum. Triangulation results of the methodological complementarity of data analysis and the search for the presence of death in the curriculum through a broad series of associated terms and codes.

The study was approved by the Research Ethics Committee of the university that coordinated the research.

Contexts of analysis

The regulations object of the study included organic laws in effect, Royal Decrees on curriculum, and other complementary state regulations applicable to education on the different levels of reference: Pre-school, Primary and Secondary education (Compulsory and High School):

- Organic Law 2/2006, dated 3 May, on Education (LOE), amended by Organic Law 8/2013, dated 9 December, on the Improvement of the Quality of Education (LOMCE).
- Royal Decree 1630/2006, dated 29 December, establishing the minimum teachings of the second cycle of pre-school education.
- Order ECI/ 3960/2007, dated 19 December, establishing the curriculum and regulating the pre-school education system.
- Royal Decree 126/2014, dated 28 February, establishing the basic curriculum for primary education.
- Royal Decree 1105/2014, dated 26 December, establishing the basic curriculum for secondary compulsory education and high school.
- Amendment of errors of Royal Decree 1105/2014, dated 26 December, establishing the basic curriculum for secondary compulsory education and high school.

Received: 04-11-2018 Accepted: 12-04-2019

- Order ECD/65/2015, dated 21 January, describing the relationships between competencies, contents and evaluation criteria of Primary, Secondary Compulsory and High School education.

The bibliographical source used for documentary consultation was the website of the Ministry of Education and Vocational Training, from which we downloaded the Official State Gazette in .pdf format. Before analysing said regulations, the contexts of analysis were assessed and validated by an expert, an inspector from the Education Department of the Autonomous Community of Madrid.

Units of analysis

After selecting and validating the primary documents, we identified the terms that were semantically linked with death awareness.

We sought the aid of 17 experts selected according to the following criteria: a minimum of 5 years of research on the Pedagogy of death, pedagogical training and work experience as teachers or tutors. The terms were analysed using a triple filter: the first based on quantitative criteria and the other two on qualitative criteria.

For this purpose, the group of experts was divided into two subgroups: A, with 12 members, and B, with 5 members. In the first filter, subgroup A was given the following instructions: "Indicate between ten and fifteen relevant terms related with the inclusion of death in education, both on the level of didactic or standard teaching materials, as well as from the tutorial perspective when a student suffers a significant loss". Once these terms were summarised, a table-summary was drafted that included the terms proposed with a frequency equal to, or greater than, 3. This task resulted in 16 terms.

For the second filter, also with subgroup A, a fifteen-day gap was given to facilitate distancing from and awareness of the proposals generated. Within a focus group framework, the group was again requested to think of any other relevant term that had not already been included. The second set of instructions was: "Which other relevant terms related with the inclusion of death in education have been omitted and could be included?". The table was completed with another 5 terms.

Received: 04-11-2018 Accepted: 12-04-2019

The third filter was comprised of presenting the results of the experts of subgroup A to those of subgroup B, who were given, also within a focus group framework, the same instructions of the second filter. Another 4 terms were generated.

The final result comprised 25 terms; the lexical roots of each were defined and used as codes for the study (Table I):

TABLE I. Terms selected for the study

Term	Lexical Root - Code	Selected in the first filter	Selected in the second filter	Selected in the third filter
Muerte [Death]	Muer.			
Duelo [Mourning]	Dol.			
Pérdida [Loss]	Perd.			
Suicidio [Suicide]	Suicid.			
Emociones [Emotions]	Emo.			
Vida [Life]	Vi.			
Fallecimiento [Passing]	Fallec.			
Despedida [Parting]	Despedi.			
Recuerdo [Memory]	Rec.			
Tabú [Taboo]	Tabú.			
Acompañamiento [Accompaniment]	Compañ.			
Mortalidad [Mortality]	Mort.			
Resiliencia [Resilience]	Resilien.			
Conciencia [Awareness]	Concien.			
Finitud [Finitude]	Fin.			
Biodiversidad [Biodiversity]	Divers.			
Salud [Health]	Salud.			
Valores [Values]	Valor.			
Tradición [Tradition]	Tradicion.			
Holocausto [Holocaust]	Holocausto.			
Existencial [Existential]	Exist.			
Ritual [Ritual]	Rit.			
Extinción [Extinction]	Extin.			
Guerra [War]	Guerr.			
Genocidio [Genocide]	Genocidio.			

Source: Prepared by the authors

Received: 04-11-2018 Accepted: 12-04-2019

Procedure

After selecting the 25 terms, we considered the presence of each code in all of the regulations, and the existence, or not, of a semantic relationship with the corresponding term. Using NVivo 12 Pro software, we entered the codes in the regulations object of analysis and implemented the following tasks:

- Quantification of the frequency in which each term appeared that, given its significance and meaning, was related with the presence of death in the curriculum, in relation to the total of the presence of the corresponding code.
- Qualitative analysis of the contents and interpretation of each term and on a global level, to compare the results within the framework of existing literature on including death in the curriculum. This analysis considered the presence and absence alike (relative or absolute) of some of the terms.

To discriminate whether or not the presence of each term or associated code was related with death awareness, we used the criteria of "fulfilment of the requirement that the term was associated with death, loss and finitude as educational elements" as a basis.

For purposes of clarification as regards this criterion, we offer an example with the term "Finitude". Within the framework of the analysis of "Earth and Environmental Sciences" in the second year of high school, it was considered pertinent for having as its "main axis the use we make as humans of our planet's resources, a *finite* planet that we 'use' as if it were unlimited" (Royal Decree 1105/2014, p. 461). It was considered not pertinent to the study when, in "Mathematics II", the code "Fin." appears in the following paragraph, with an irrelevant significance and meaning: "Limit of a function at a point and in *infinitude*" (Royal Decree 1105/2014, p. 421).

Results

Quantitative analysis

An analysis is displayed of the frequencies of the presence of the 25 terms –based on their corresponding codes– by education level (Table II),

Received: 04-11-2018 Accepted: 12-04-2019

general reference regulation (Table III), and by areas of knowledge (Table IV). The total number of appearances of each term, the number of times in which the presence relates to death awareness, and the percentage of the relationship in reference to the total number of appearances, are included.

TABLE II. Quantitative analysis of terms and their relationship with death awareness by education level

	Pre-school education			**Primary education**			**Secondary education, compulsory and high school**		
	N total	**N relationship**	**% relationship**	**N total**	**N relationship**	**N total**	**N total**	**N relationship**	**% relationship**
Acompañamiento [Accompaniment]	0	0	0%	0	0	0%	2	0	0%
Biodiversidad [Biodiversity]	0	0	0%	0	0	0%	32	7	21.87%
Conciencia [Awareness]	24	0	0%	12	0	0%	68	1	1.47%
Despedida [Parting]	0	0	0%	6	0	0%	1	0	0%
Duelo [Mourning]	2	0	0%	0	0	0%	1	0	0%
Emociones [Emotions]	48	1	2.08%	14	0	0%	64	0	0%
Existencial [Existential]	0	0	0%	0	0	0%	4	0	0%
Extinción [Extinction]	0	0	0%	1	1	100%	5	5	100%
Fallecimiento [Passing]	0	0	0%	0	0	0%	0	0	0%
Finitud [Finitude]	0	0	0%	0	0	0%	19	1	5.26%
Genocidio [Genocide]	0	0	0%	0	0	0%	1	1	100%
Guerra [War]	0	0	0%	0	0	0%	117	117	100%
Holocausto [Holocaust]	0	0	0%	2	2	100%	7	7	100%
Mortalidad [Mortality]	0	0	0%	0	0	0%	5	5	100%
Muerte [Death]	2	2	100%	0	0	0%	12	12	100%
Pérdida [Loss]	0	0	0%	0	0	0%	21	19	90.47%
Recuerdo [Memory]	2	0	0%	0	0	0%	0	0	0%

Received: 04-11-2018 Accepted: 12-04-2019

Resiliencia [Resilience]	0	0	0%	1	1	100%	0	0	0%
Ritual [Ritual]	0	0	0%	0	0	0%	9	0	0%
Salud [Health]	21	2	9.52%	31	15	48.38%	115	33	28.69%
Suicidio [Suicide]	0	0	0%	0	0	0%	0	0	0%
Tabú [Taboo]	0	0	0%	0	0	0%	7	3	47.85%
Tradición [Tradition]	6	0	0%	7	0	0%	53	0	0%
Valores [Values]	11	0	0%	81	1	1.23%	305	2	0.65%
Vida [Life]	69	2	2.89%	92	7	7.60%	378	28	10.05%

Source: Prepared by the authors

TABLE III. Quantitative analysis of terms and their relationship with death awareness by general education regulation

	LOE			LOMCE			ORDER OF KEY COMPETENCIES		
	N total	N relationship	% relationship	N total	N relationship	% relationship	N total	N relationship	% relationship
Acompañamiento [Accompaniment]	0	0	0%	0	0	0%	0	0	0%
Biodiversidad [Biodiversity]	0	0	0%	0	0	0%	0	0	0%
Conciencia [Awareness]	0	0	0%	0	0	0%	0	0	0%
Despedida [Parting]	0	0	0%	0	0	0%	0	0	0%
Duelo [Mourning]	0	0	0%	0	0	0%	0	0	0%
Emociones [Emotions]	2	0	0%	1	0	0%	3	0	0%
Existencial [Existential]	0	0	0%	0	0	0%	0	0	0%
Extinción [Extinction]	2	0	0%	2	0	0%	0	0	0%
Fallecimiento [Passing]	0	0	0%	0	0	0%	0	0	0%
Finitud [Finitude]	0	0	0%	0	0	0%	0	0	0%
Genocidio [Genocide]	0	0	0%	0	0	0%	0	0	0%
Guerra [War]	0	0	0%	0	0	0%	0	0	0%
Holocausto [Holocaust]	1	1	100%	1	1	100%	0	0	0%

Received: 04-11-2018 Accepted: 12-04-2019

Mortalidad [Mortality]	0	0	0%	0	0	0%	0	0	0%
Muerte [Death]	0	0	0%	0	0	0%	0	0	0%
Pérdida [Loss]	0	0	0%	0	0	0%	0	0	0%
Recuerdo [Memory]	0	0	0%	0	0	0%	0	0	0%
Resiliencia [Resilience]	0	0	0%	0	0	0%	0	0	0%
Ritual [Ritual]	0	0	0%	0	0	0%	0	0	0%
Salud [Health]	8	1	12.5%	2	1	50%	4	0	0%
Suicidio [Suicide]	0	0	0%	0	0	0%	0	0	0%
Tabú [Taboo]	0	0	0%	0	0	0%	0	0	0%
Tradición [Tradition]	1	0	0%	0	0	0%	1	0	0%
Valores [Values]	44	1	2.27%	23	2	8.69	25	0	0%
Vida [Life]	40	0	0%	17	0	0%	25	0	0%

Source: Prepared by the authors

TABLE IV. Quantitative analysis of terms and their relationship with death awareness by area of knowledge in the education levels

Term	Education level	Areas of knowledge	N (relationship)
Biodiversidad [Biodiversity]	Secondary	Biology and Geology	4
		Earth and Environmental Sciences	1
		Scientific Culture	2
Conciencia [Awareness]	Secondary	Introduction to Classic Culture	1
Emociones [Emotions]	Pre-school	Self-awareness and Personal Autonomy	1
Extinción [Extinction]	Primary	Natural Sciences	1
	Secondary	Biology and Geology	3
		Introduction to Scientific Culture	1
Finitud [Finitude]	Secondary	Earth and Environmental Sciences	1
Genocidio [Genocide]	Secondary	Ethical Values	1
Guerra [War]	Secondary	Fundamentals of Art	32
		Geography and History	46
		History of the Contemporary World	39

Received: 04-11-2018 Accepted: 12-04-2019

Holocausto [Holocaust]	Primary	Social and Civic Values	2
	Secondary	Transversal Elements	1
		Geography and History	3
		History of the Contemporary World	2
		Ethical Values	1
	LOE and LOMCE	Prevention and peaceful conflict resolution and basic values of democracy and human rights	2
Mortalidad [Mortality]	Secondary	Fundamentals of Art	1
		Greek	1
		History of Spain	1
		History of Philosophy	1
		Classic Culture	1
Muerte [Death]	Pre-school	Knowledge of the environment	2
	Secondary	Philosophy	6
		Fundamentals of Art and Fundamentals of Art II	2
		Geography and History	1
		History of Spain	1
		History of Art	2
Pérdida [Loss]	Secondary	Biology and Geology	7
		General and Business Economics	2
		Fundamentals of Art I	1
		History of Spain	2
		History of Philosophy	2
		Scientific Culture	2
		Initiation to Entrepreneurship and Business Development	1
		Psychology	1
		Ethical Values	1
Resiliencia [Resilience]	Primary	Social and Civic Values	1

Received: 04-11-2018 Accepted: 12-04-2019

Salud [Health]	Pre-school	Self-awareness and Personal Autonomy	2
	Primary	Transversal Elements	1
		Natural Sciences	14
	Secondary	Biology and Geology	14
		Chemistry	1
		Applied Anatomy	9
		Physical education	9
	LOE and LOMCE	Objectives of Vocational Training	2
Tabú [Taboo]	Secondary	Spanish Language and Literature	3
Valores [Values]	Primary	Transversal Elements	1
	Secondary	Transversal Elements	1
		Physical education	1
	LOE and LOMCE	Purpose of the Education System	3
Vida [Life]	Pre-school	Knowledge of the environment	2
	Primary	Natural Sciences	1
		Social Sciences	5
		Social and Civic Values	1
	Secondary	Biology and Geology	11
		Biology	5
		Economics	2
		Philosophy	10
		Physics	1
		Fundamentals of Art	1
		Geography and History	4
		Geography	1
		Initiation to Entrepreneurship and Business Development	1
		History of the Contemporary World	1
		Scientific Culture	1

Source: prepared by the authors

Qualitative analysis

Then, we analysed the presence in the context of terms related with death awareness and, therefore, comparable to the Pedagogy of death. Of likewise importance is the analysis of the presences as well as the

interpretation of the absences in the curriculum of some terms and codes, detailed in greater depth in "Scientific Discussion".

The term "Biodiversity" appeared in high school as regards the loss and threat of extinction of species in subjects like "Biology and Geology" or "Earth and Environmental Sciences". One example is given in the standard of learning, "Indicate the main measures that reduce the loss of *biodiversity*" (Royal Decree 1105/2014, p. 214).

Worth mentioning is the relative absence of the term "Awareness", related with death on only one occasion, within the framework of the subject "Classic Culture", in compulsory secondary education (ESO): "To raise *awareness* of the survival, influence and presence of many of these aspects (general approach to the study of Greek and Latin civilisations) in Western culture" (Royal Decree 1105/2014, p. 468). Opposite to the case of the term "Biodiversity", the inclusion of the term "Awareness" in the curriculum is not associated with loss, but rather with survival and, therefore, with diachronic awareness.

Another term with a high frequency but scarce association with education on death awareness is "Emotions". Just one interesting reference is found for our purposes - an indirect relationship with emotions that death and loss may trigger - in pre-school education (Order ECI 3960/2007): "Increasing association and verbalisation of causes and consequences of basic *emotions*, like love, happiness, fear, sadness or anger" (p. 1021).

The term "Extinction" appears in primary and secondary education curriculum in subjects related with Natural Sciences, and in all cases referring to the extinction of species. For example, "Identify and explain some of the causes of *extinction* of a species" (Royal Decree 126/2014, p. 18) in the subject "Natural Sciences" in primary education.

On only one occasion is the term "Finitude" related with death awareness. Specifically, in the subject "Earth and Environmental Sciences" in the second year of high school (Royal Decree 1105/2014), when a description is offered in that the subject has as its "main axis the use we make as humans of our planet's resources, a *finite* planet that we "use" as if it were unlimited" (p. 461).

"Genocide" also appears once in the subject "Ethical Values" at the ESO level: "Political rights: wars, terrorism, dictatorships, *genocide*, political refugees, etc." (Royal Decree 1105/2014, p. 540).

Received: 04-11-2018 Accepted: 12-04-2019

Their semantic lineage is coherent with the terms "Genocide", "War" and "Holocaust". The presence of the first one is important in secondary education, given its contents, related with the historical analysis of inciting factors, the consequences and the development of wars, directly related with the Pedagogy of death. The term "Holocaust" is clearly present in the curriculum, especially in primary and secondary education, not only in subjects like "Social and Civic Values" or "Geography and History", but also as a transversal element of both levels. "Article 2. Transversal Elements" of Royal Decree 1105/2014, on secondary education, cites:

> The teaching programme must comprise, in any case, prevention of gender violence, violence against persons with disabilities, terrorism and any other type of violence, racism or xenophobia, including the study of the Jewish *Holocaust* as a historical occurrence (p. 174).

The term "Mortality" appears in the study of religions, myths and rituals, in secondary education. For example, in the subject "Fundamentals of Art" in the first year of high school, "Worship of the dead, *immortality* and resurrection" in ancient Egypt is introduced as content (Royal Decree 1105/2014, p. 281). It is also touched upon through History in the demographic study of certain eras marked by a "high *mortality*" (Royal Decree 1105/2014, p. 325).

The term "Death" appears in both pre-school and secondary education. In pre-school, its presence is contextualised within "Knowledge of the environment" (Royal Decree 1630/2006): "Observation of characteristics, behaviours, functions and changes in living beings. Consideration of the life cycle, birth and *death*" (p. 10). Worth mentioning is its disappearance in primary education, and its return in secondary education. It does so through the subjects: (a) "Philosophy", for example, as content "The issue of meaning, essence and existence, the self, freedom, *death*, destiny, chance, History, the need for significance" (Royal Decree 1105/2014, p. 254); (b) "Fundamentals of Art I and II", with standards of learning such as "Analyse the painting `The *death* of Sardanapalus´" (Royal Decree 1105/2014, p. 288); (c) "Geography and History", relating present and past: "Acknowledging that the past `is neither *dead* nor buried´, but rather determines or impacts our present and different possible futures and different spaces" (Royal Decree 1105/2014, p. 304); (d) "History of Spain", through the study of important deaths during the country's

historical development: "Explain the political alternatives proposed after the *death* of Franco" (Royal Decree 1105/2014, p. 327); and (e) "History of Art", for the presence of death in art: "Identify, analyse and comments the following paintings of the 17th-century European Baroque: *Calling of Saint Matthew* and *Death of the Virgin*, by Caravaggio" (Royal Decree 1105/2014, p. 338).

The term "Loss" is characterised for its link to death awareness only in the secondary education curriculum. Loss is associated with biodiversity and natural resources in "Biology and Geology", and also with loss and death from a historical perspective. For example, in regards to the "*loss* of hegemony in Europe" (Royal Decree 1105/2014, p. 323), it indicates the end of a historical era and, therefore, educates on awareness of finitude. From the dialectics of the individual and society, "loss" is also mentioned when including, in "History of Philosophy", the "*loss* of the individual in the face of mass culture" (Royal Decree 1105/2014, p. 333).

"Resilience" appears just once in the curriculum. Specifically, in the regulatory framework of primary education (Royal Decree 126/2014), in the subject "Social and Civic Values": "Describe the value of cognitive restructuring and *resilience*" (p. 52).

The term "Health" has been included in the study when it refers to accident prevention or triggering factors of illness and, when applicable, death. Its appearance in the curriculum entails this meaning across all education levels. For example, in pre-school, it is framed within the area "Self-awareness and Personal Autonomy" (Royal Decree 1630/2006):

> Adequate assessment of risk factors, adoption of preventive and safety behaviours in everyday situations, calm and collaborative attitude in situations of illness and minor accidents. Identification and critical assessment when facing everyday factors and social practices that are favourable or unfavourable to *health* (p. 8).

In many social, cultural and educational contexts, death, from the perspective of its meaning and as an action of communication itself alike, remains a taboo concept and term. In secondary education curriculum, the term "Taboo" indirectly relates with death, in contents like the "Observation, reflection and explanation of changes that affect the meaning of words: causes and mechanisms. Metaphor, metonymy, *taboo* words and euphemism" (Royal Decree 1105/2014, p. 363), in this case in the subject "Spanish Language and Literature".

Received: 04-11-2018 Accepted: 12-04-2019

We must highlight the presence and list of values related with loss and death, among the purposes of education. In this regard, the term "Values" is expressed as a transversal factor in primary and secondary education. For example, the following citation applicable to primary education (Royal Decree 126/2014):

> Values that are the basis of freedom, justice, equality, political pluralism, peace, democracy, respect for human rights and the rejection of terrorism, plurality, respect for the rule of law, respect and consideration for victims of terrorism and the prevention of terrorism and any type of violence. In any case, the teaching programme must comprise violence prevention (p. 8).

Among educational goals included in the LOE and LOMCE, references are also made to values that avoid violent death and promote respect for human rights. The LOMCE specifies that the education system must promote "*Values* that are pillars of democracy and human rights, including, in any case, the prevention of gender violence and the study of the Jewish holocaust as a historical occurrence" (p. 295).

Finally, as regards the term "Life", it relates to the inclusion of death in the curriculum in all three education levels. In pre-school education, in addressing life through the perspective of an interdependent ecosystem in which the life cycle, the relationship between living beings and inert matter comprises the vital space of survivors. In primary education, the concept of life expectancy is first introduced in the subject "Social Sciences". In secondary education, the contents of preceding learning levels are updated and death awareness is introduced in subjects like "Economy" -for example, in the content of "Planning the future. Economic needs in the stages of *life*" (Royal Decree 1105, 2014, p. 243)- or "Philosophy" (Royal Decree 1005, 2014):

> Reflect on the question of the meaning of existence, explaining the central theses of some philosophical theories about *life*, and present a reflective discourse on life or death, or the historical path, or the role of the individual in reality, among other metaphysical issues (p. 494).

Received: 04-11-2018 Accepted: 12-04-2019

Scientific Discussion

The scientific discussion is based on the following elements: (a) Presence in areas of knowledge and subjects; (b) Presence by education levels; (c) Presence by Pedagogy of death approach; and (d) Relevant absences.

We may conclude from the documentary study that the subjects (a) that most often include curricular elements that may relate with the Pedagogy of death are environmental in focus. For example, "Natural Sciences" and "Earth and Environmental Sciences", that draw the learner closer to the importance of nature. In greater depth, the subject "Biology and Geology" focuses on the critical analysis of ecosystems, the importance and causes of biodiversity and the extinction of species, delving into the consequences and significance of the loss.

Other types of subjects highly related with death awareness are comparable to Social Sciences, Humanities and the Arts. These underscore the teaching and learning of the significance of events that have shaped our present, encompassing the relationship between past-present-future and generating a historical awareness of finitude. Worth mentioning are wars, the Holocaust or humanitarian crises, confirming, in line with the contributions of other authors (Bos, 2014; Lindquist, 2007; Mariana, 2016; Zembylas, 2011) that these contents may be of extraordinary interest in the development of death awareness. In any case, what stands out is the clear presence of teaching about the "Holocaust", in subjects and also as a transversal element of the curriculum, in comparison with the presence of "Genocide". Our opinion is that a conscious and planned Pedagogy of death must teach about the Holocaust but without overlooking or omitting other crimes against humanity of lesser social visibility.

Also, death awareness is addressed in philosophy, given its promotion of an awareness of finitude, the meaning of life and, therefore, of death, the human being, reality and metaphysics, practical reasoning, etc. This interpretation of the Pedagogy of death as a starting point for educating on existential competencies coincides with previous studies of other authors (i.e., Mantegazza, 2004; Mèlich, 1989).

Death is also explicitly addressed through other subjects, like "Fundamentals of Art" and "Classic Culture", which teach about major cultures, their traditions and main rites (many of these associated with behaviours following death).

Received: 04-11-2018 Accepted: 12-04-2019

The subject "Spanish Language and Literature" during secondary education includes the taboo as object of study through its semantic analysis. Transforming death into myth and taboo continues to be one of the main handicaps to achieving a true Pedagogy of awareness. However, subjects like "Spanish Language and Literature" comprise areas which may include death as one of postmodern society's main taboos.

Finally, subjects like "Ethical Values" and "Social and Civic Values" are examples of how to educate on death awareness, by promoting values for the sake of a humanised society and by respecting the Universal Declaration of Human Rights and international treaties in effect based on it.

By education levels (b), we may point out the existence of curricular elements ascribed to pre-school, primary and secondary education levels that are likely to become, through organised and strategic planning, part of a Pedagogy of death applicable to any level (Herrán and Cortina, 2006; Jackson and Colwell, 2002). However, the interpretation of the results allows us to observe that the presence is greater and more clearly linked to death awareness in pre-school and secondary education levels. In pre-school, death awareness is addressed through the area "Knowledge of the environment" in relation to the life cycle. In secondary education, as observed, it is addressed through different subjects. Its presence is more tangential in primary education, through the introduction of related curricular elements, mainly associated with health education, or in considering the Holocaust as transversal content.

Given the focus (c) upon which we may propose the Pedagogy of death -preliminary and palliative- (Cortina and Herrán, 201; Herrán and Cortina, 2006; Herrán et ál., 2000), we can affirm that death is present in curriculum design from pre-school to secondary education, though in an unplanned, non-strategic way. In fact, neither the principles and purposes of education defined in the LOE and LOMCE, nor in the objective of each level established in their corresponding regulations, do we find any mention whatsoever of promoting death awareness in education. As regards the palliative approach of accompaniment in situations of grief, no mention or consideration is made, for example, of its susceptibility as a specific need of students in terms of educational support. In fact, as to the terms "Grief" or "Passing", we find n = 0 as regards a direct relationship with loss and death.

Other relevant absences (d) are worth mentioning, like "Suicide" or "Parting". Suicide is the leading cause of death among adolescents in European countries (World Health Organization, 2017). However, it is overlooked by regulations applicable to the education system. As to "Parting", it is absent in primary education, despite being essential content in primary education (Geddes, 2006) and its consideration as the starting point of a Pedagogy of death, which may be helpful for establishing healthy emotional ties. Also worth emphasising is the scarce relevance of death in the field of competencies-based education, given the complete absence (n = 0) of terms related with death in "Order ECD/65/2015, dated 21 January", describing the 7 key competencies of the Spanish education system. The authors of this manuscript uphold the idea that an education based on competencies (Valle and Manso, 2013) that fails to educate on what is most important for human beings, highlighting death -both one's own and that of others-, remains an incomplete education that is incoherent with the human condition itself and that cannot sufficiently prepare students for life.

Conclusions, limitations and foresight

The analysis of terms related with death awareness present in the curriculum depicts a rigorous map of the current status of curricular design as regards death in the Spanish education system. Likewise, its presence defines both learning opportunities and gaps and limitations for educating on death.

Globally, we may appreciate that in those areas in which death-related terms appear, the relevant contents are actually others. In other words, death is secondary content that is included merely tangentially within other areas given higher priority. This includes the LOE and LOMCE texts alike and the curriculum of the subjects object of the study. Several conclusions may be deduced from this observation:

- That death and, thereby, the Pedagogy of death, are not prioritised in the texts analysed.
- That the presence of death responds, to date, to its transversal nature, despite its failure to bring together the features of transversal themes or elements, wherefore it has been categorised as a "radical issue" (Herrán et ál., 2000).

Received: 04-11-2018 Accepted: 12-04-2019

- That, given its presence in life and in human relationships, it is impossible to overlook death in the curriculum, even if weakly and indirectly.
- That a profoundly enriching horizon for research and learning opens up at the heart of education and the emerging society.
- That the absence of a greater presence may still reflect its taboo nature, the lack of reflection or inquiry on its educability or the prioritisation of other fields and contents.
- That the usual focus of education seems to suffer from epistemic short-sightedness, polarised on what's close-up, contextual, immediate, demanded and excluded, *de facto*, not only in such a transcendental issue as death, but also in relation to other teaching-related challenges that share the common factor of education on awareness.
- That the potential exists for greater educability of death, through a more direct and conscious approach at all education levels, at least through those subjects that include relevant terms and contents.
- That, to achieve genuine education of human beings, the curriculum must be modified to include death as a reference of learning.

Considering the deductive logic of the curriculum to be designed as of the principles and purposes of education, the absence of death awareness in these makes impossible for planning and developing an adequate curriculum of the Pedagogy of death. Therefore, this study's results and conclusions are the basis for proposing the goal of the Spanish education system in this regard: "Education on death awareness, adapted to each age and stage of education, in such manner that it may contribute to students an education in values that takes into account the human condition as its main vector, humanity as an essential identity and compatible with other cultural identities, promoting personal and social maturity and sociohistorical responsibility".

This study presents some limitations. The first, without a doubt, is that the 25 selected terms do not respond to the exhaustiveness with which death may appear in the curriculum. This delimitation arises from the required succinctness to guarantee the study's viability. The second relates to the contexts of analysis, which have been restricted to an analysis of state regulations, excluding the subsequent development of the curriculum by the autonomous communities.

Given this limitation, we deduce that the application of the study to the curriculum defined on the level of autonomous communities and the corresponding comparative analysis may entail a future line of research. Another line of research is the analysis of the presence of death awareness in key documents of supranational bodies, such as the European Union, the UNESCO (United Nations Educational, Scientific and Cultural Organization), the OEI (Organisation of Ibero-American States for Education, Science and Culture), the ALECSO (Arab League Educational, Cultural and Scientific Organization), etc.

All of these will contribute toward drawing a complete map of the consideration of death in education, a necessary diagnosis for finding key references and significant absences alike of an issue that is as human and educational as death.

Bibliographic references

Aspinall, S.Y. (1996). Educating children to cope with death: a preventive model. *Psychology in the Schools, 33*(4), 341-349.

Berg, C.D. (1982). Helping children accept death and dying through group counseling. *Personal and Guidance Journal*, *57*(3), 169-172.

Bos, P. R. (2014). Empathy, sympathy, simulation? Resisting a holocaust pedagogy of identification. *Review of Education, Pedagogy, and Cultural Studies*, *26*(5), 403-421.

Colomo, E. (2016). Pedagogía de la muerte y proceso de duelo. Cuentos como recurso didáctico. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, *14*(2), 63-77.

Colomo, E. and Oña, J. M. de (2014). Pedagogía de la muerte. Las canciones como recurso didáctico. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, *13*(3), 109-121.

Corr, C., Nabe, C., and Corr, D. (2000). *Death and dying: life and living*. California: Wadsworth Pub.

Corrección de errores del Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la educación secundaria obligatoria y del bachillerato. Boletín Oficial del Estado 1/05/2015. [Amendment of errors of Royal Decree 1105/2014, dated 26 December,

Received: 04-11-2018 Accepted: 12-04-2019

establishing the basic curriculum for secondary compulsory education and high school].

Cortina, M. and Herrán, A. de la (2011). *Pedagogía de la muerte a través del cine*. Madrid: Universitas.

Edgar, L. and Howard-Hamilton, M. (1994). Non-crisis death education in the elementary school. *Guidance and Counseling, 29*(1), 38-46.

Fertziger, A.P. (1983). Right- and left- brain approaches to death education. *Death Education*, *7*(1), 33-38.

Fullat, O. (1982). *Las finalidades educativas en tiempos de crisis*. Barcelona: Hogar del Libro.

Geddes, H. (2006). *El apego en el aula. Relación entre las primeras experiencias infantiles, el bienestar emocional y el rendimiento escolar*. Barcelona: Editorial Graó.

Glass, J. (1990). Changing death anxiety through death education in the public schools. *Death Studies, 14*(1), 31-52.

Herrán, A. de la and Cortina, M. (2006). *La muerte y su didáctica. Manual para educación infantil, primaria y secundaria*. Madrid: Universitas.

(2008). La práctica del acompañamiento educativo desde la tutoría en situaciones de duelo. *Tendencias Pedagógicas*, 13, 157-173.

Herrán, A. de la, González, I., Navarro, M. J., Freire, M. V. and Bravo, S. (1998). *La educación para la muerte como tema transversal de transversales*. World Congress of Early Childhood Education, Madrid, 26 December.

(2000). *¿Todos los caracoles se mueren siempre? Como tratar la muerte en educación infantil*. Madrid: De la Torre.

Hetzel, S., Winn, V., and Tolstoshev, H. (1991). Loss and change: new directions in death education for adolescents. *Journal of Adolescence, 14*(4), 323-334.

Hulbert, F. (1999). Teaching about loss, change, death and bereavement in the primary school. *The Journal of Progressive Judaism*, *13*(13), 79-92.

Jackson, M., and Colwell, J. (2002). *A teacher´s handbook of death*. London: Jessica Kingsley Publishers.

Jaramillo, J. (2017). *Educación para la muerte: imaginarios sociales del docente y del estudiante universitario en Colombia* (Tesis inédita de Doctorado). Universidad Autónoma de Madrid, Madrid.

Kroen, W.C. (2002). *Cómo ayudar a los niños a afrontar la pérdida de un ser querido. Un manual para adultos*. Barcelona: Oniro.

Ley Orgánica 2/2006, de 3 de mayo, de Educación (LOE). Boletín Oficial del Estado 4/5/2006. [Organic Law 2/2006, dated 3 May, on Education (LOE)].

Ley Orgánica 8/2013, de 9 de diciembre, para la Mejora de la Calidad Educativa (LOMCE). Boletín Oficial del Estado 10/12/2013. [Organic Law 8/2013, dated 9 December, on the Improvement of the Quality of Education (LOMCE)].

Lindquist, D. H. (2007). Avoiding inappropriate pedagogy in middle school teaching of the holocaust. *Middle School Journal*, *39*, 24-31.

Mantegazza, R. (2004). *Pedagogia della morte*. Milano: Città Aperta.

Mariana, M. L. (2016). *El holocausto en la escuela desde la Orientación Educativa: Fundamentos para una Pedagogía de la Shoá* (Tesis inédita de Doctorado). Universidad Autónoma de Madrid, Madrid.

Mèlich, J.-C. (1989). *Situaciones-límite y educación. Estudio sobre el problema de las finalidades educativas.* Barcelona: PPU.

Orden ECD/65/2015, de 21 de enero, por la que se describen las relaciones entre las competencias, los contenidos y los criterios de evaluación de la educación primaria, la educación secundaria obligatoria y el bachillerato. Boletín Oficial del Estado 29/01/2015. [Order ECD/65/2015, dated 21 January, describing the relationships between competencies, contents and evaluation criteria of Primary, Secondary Compulsory and High School education].

Orden ECI/3960/2007, de 19 de diciembre, por la que se establece el currículo y se regula la ordenación de la educación infantil. Boletín Oficial del Estado 5/01/2008. [Order ECI/ 3960/2007, dated 19 December, establishing the curriculum and regulating the pre-school education system].

Payne, G. and Payne, J. (2004). *Key concepts in social research*. London: Sage Publications.

Petitfils, B. (2016). Encountering mortality: A decade later, the pedagogical necessity of Six Feet Under. *International Journal of Pedagogy and Curriculum*, *13*(3), 249-260.

Pine, V.R. (1977). A sociohistorical portrait of death education. *Death Education*, *1*(1), 57-84.

Real Decreto 1630/2006, de 29 de diciembre, por el que se establecen las enseñanzas mínimas del segundo ciclo de educación infantil. Boletín Oficial del Estado 4/01/2007. [Royal Decree 1630/2006, dated 29 December, establishing the minimum teachings of the second cycle of pre-school education].

Received: 04-11-2018 Accepted: 12-04-2019

Real Decreto 126/2014, de 28 de febrero, por el que se establece el currículo básico de la educación primaria. Boletín Oficial del Estado 1/03/2014. [Royal Decree 126/2014, dated 28 February, establishing the basic curriculum for primary education].

Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la educación secundaria obligatoria y del bachillerato. Boletín Oficial del Estado 3/01/2015. [Royal Decree 1105/2014, dated 26 December, establishing the basic curriculum for secondary compulsory education and high school].

Rodríguez, P., Herrán, A. de la and Cortina, M. (2012). Antecedentes de la Pedagogía de la muerte en España. *Enseñanza & Teaching*, 30(2), 175-195. (2015a).

(2015a). Pedagogía de la muerte mediante aprendizaje de servicio. *Educación XX1*, *18*(1), 189-212.

(2015b). *Educar y vivir teniendo en cuenta la muerte*. Madrid: Pirámide.

(2019). Antecedentes internacionales de la Pedagogía de la muerte. *Foro de Educación, 17*(26), 259-276.

Sahler, O. (1983). *El niño y la muerte.* Madrid: Alhambra.

Shackelford, T. K. (2003). Assessing individual differences in death threat: A brief theoretical and psychometric review of the Threat Index. *Omega: Journal of Death and Dying, 46*, 323-333.

Turner, M. (2004). *Cómo hablar con niños y jóvenes sobre la muerte y el duelo.* Barcelona: Paidós.

Valle, J. M. y Manso, J. (2013). Competencias clave como tendencia de la política educativa supranacional de la Unión Europea. *Revista de Educación*, *Extraordinario 2013*, 12-33.

Verdú, V. (2002). La enseñanza del fin. *El País*, 5 de julio.

World Health Organization (2017). *Global accelerated action for the health of adolescents (AA-HA!).* Paris: WHO.

Zembylas, M. (2011). Personal narratives of loss and the exhumation of missing persons in the aftermath of war: In search of public and school pedagogies of mourning. *International Journal of Qualitative Studies in Education*, *24*(7), 767-784.

Contact address: Agustín de la Herrán Gascón. Universidad Autónoma de Madrid, Facultad de Formación de Profesorado y Educación, Departamento de Didáctica y Teoría de la Educación. C/ Francisco Tomás y Valiente, 3, 28049 Madrid. E-mail: agustin.delaherran@uam.es

Received: 04-11-2018 Accepted: 12-04-2019

Assessment of Reading Comprehension in a Diglossic Context

Evaluación de la comprensión lectora en un contexto diglósico

DOI: 10.4438/1988-592X-RE-2019-385-423

Ana M. Rico-Martín
Universidad de Granada
Amaruch Mohamedi-Amaruch
CEIP España

Abstract

Reading comprehension is an essential ability for students to cope efficiently in and out of the classroom, but it is a complex skill that does not always guarantee that the strategies generated in the mother tongue are extrapolated into a person's second language because of the differences between both languages, as is the case with the bilingual students participating in this research.

In this study, we analyse the differences that exist in the reading comprehension of 612 students finishing Primary Education (Level 6) in the Autonomous City of Melilla (Spain), divided into two groups: one monolingual (Spanish) and the other bilingual (Berber-Spanish), from the perspective of the mother tongue and the language of education (Spanish), in a Spanish city characterised by diglossia among much of its population and a submersion educational model, but with attention to the culture of origin.

Having carried out the ACL-6 test (*Avaluació de la Comprensió Lectora 6* [Reading Comprehension Assessment 6]) (Catalá, Català, Molina & Monclús, 2008) in this empirical-analytical research, with an *ex-post-facto* descriptive study and the use of descriptive and inferential statistics, the results show the general low level of the students in all reading dimensions (literal, reorganizational, inferential and critical) and according to the textual typology used (literary –narrative and poetic–, expository and interpretation of data texts –with graphics and tables–),

Received: 29-06-2018 Accepted: 26-04-2019

almost 70% of the students present a degree of reading comprehension below the average expected for their school year (particularly in the reorganizational dimension and in the expository texts). But this level proves to be even lower in all analysed variables in the case of the bilingual group.

Keywords: Reading Comprehension, Reading Processes, Bilingual Students, Spanish as a Second Language, Diglossic Context, Primary Education.

Resumen

La comprensión lectora es una habilidad fundamental para que el alumnado pueda desenvolverse eficientemente tanto dentro como fuera del aula, pero es una destreza compleja que no siempre garantiza que las estrategias generadas en la lengua materna se extrapolen hacia la segunda lengua de una persona debido a las diferencias entre ambos idiomas, como es el caso del alumnado bilingüe participante en esta investigación.

En este trabajo se analizan las diferencias existentes en la comprensión lectora de 612 estudiantes que acaban la Educación Primaria (6.º curso) en la Ciudad Autónoma de Melilla (España), distribuidos en dos grupos: uno monolingüe (español) y otro bilingüe (bereber-español), desde la perspectiva de la lengua materna y la lengua vehicular de la enseñanza (español), en una ciudad española caracterizada por una diglosia en gran parte de su población y un modelo educativo de sumersión, pero con atención a la cultura de origen.

Tras la implementación de la prueba ACL-6 (*Avaluació de la Comprensió Lectora 6*) (Catalá, Català, Molina y Monclús, 2008) en esta investigación empírico-analítica, con un estudio descriptivo *ex-post-facto* y el empleo de estadísticos descriptivos e inferenciales, los resultados señalan el bajo nivel general del alumnado en todas las dimensiones lectoras (literal, de reorganización de la información, inferencial y crítica) y según la tipología textual empleada (textos literarios –narrativos y poéticos–, expositivos y de interpretación de datos –con gráficos y tablas–), casi el 70% de los estudiantes presentan un grado de comprensión lectora inferior al promedio esperado para su nivel académico (sobre todo en la dimensión de reorganización de la información y en los textos expositivos), pero aún es más bajo en todas las variables analizadas en el caso del grupo bilingüe.

Palabras clave: comprensión lectora, procesos lectores, alumnado bilingüe, español como segunda lengua, contexto diglósico, Educación Primaria.

Introduction

Reading competence is a determining factor for the academic and professional future of a person, a low level of reading comprehension

makes it difficult to acquire the curricular content of the subjects of any educational stage. This likewise makes it difficult to attain the social and civic profile required by society today.

In other words, reading involves many necessary skills for personal and social development, and for integration and promotion in the labour market so that active participation in society is facilitated (Eurydice, 2011; Kendeou, van den Broek, Helder, & Karlsson, 2014; Solé, 2012). This importance of reading justifies this competence having a prominent place in national and international tests (such as PISA, *Programme for International Student Assessment*, and PIRLS, *Progress in International Reading Literacy Study*). The 2009 PISA report (Instituto de Evaluación [Institute of Evaluation], IE, 2010a) is the last that has results from the city of Melilla, where this study is based. In it, the schools of Melilla obtained the worst results both in Spain and compared to other nations (Consejo Escolar del Estado [State School Council], 2016).

The purpose of this study is to examine the various levels of reading that lead to the global comprehension of a text, in order to then analyse the existing differences between two groups of readers, one monolingual and the other bilingual, from the perspective of the first language (L1) and the language used in primary education. This is carried out in a Spanish city that is characterized by the diglossia with bilingualism of a sizeable proportion of its population. Moreover, there are few studies on the linguistic situation of Berber-Spanish bilinguals (Herrera & Defior, 2005; Herrera, Defior, Serrano, & Jiménez-Fernández, 2009; Molina, 2015; Olmos, 2009).

Reading Processes

Kintsch (2003) links the process of reading comprehension to the mental representation of a text from a network of interrelated propositions, established from the reading of the text itself, along with the background knowledge and personal experience of the reader. Similarly, there are authors who show that this three-way connection (textual information-background knowledge-reader goals) gives rise to the interaction of various processes necessary for understanding a text when read (Goldamn, McCarthy, & Burkett, 2015; Kendeou *et ál.*, 2014; van den Broek & Helder, 2017).

These are gradual thought processes (dimensions of reading) that, despite dating back to Barrett's taxonomy (1968), are those analysed in the different Spanish educational assessment tests (Instituto Nacional de Evaluación Educativa [National Institute of Educational Assessment], INEE, 2012), to which the critical or evaluative dimension has been added (see Table I).

TABLE I. Characteristics of the reading dimensions

Dimension	Characteristics	Strategies
Literal	Corresponds to comprehension of what is explicit in the text. It is the lowest level as it simply shows the superficial part of the text's information.	Distinguish between relevant and secondary information. Master the basic vocabulary…
Information reorganization	It is reached when the reader is capable of establishing logical connections between ideas and expressing them in a different way.	Organize, analyse and summarize the information of the text in a different way.
Inferential/interpretative	The reader goes beyond the direct meaning of the passage. Background knowledge is activated and hypotheses are formulated about the content of the text from what is given, which are verified or reformulated while reading.	Infer details, main ideas, sequences… Establish relations of cause and effect, character traits…
Critical/evaluative	It requires a process of evaluation and judgement by the reader about the ideas read, according to their knowledge and experience, to subsequently express personal opinions about their reading.	Compare the ideas from the text with criteria that are external (teacher, authorities) or internal (own experience). Determine the intentions of the text's author. Make judgements on the reality, on facts and opinions.

Source: authors.

Regarding these dimensions, the interest in students' reading comprehension has given rise to many studies similar to the research reflected here, but with results varying according to the context: Carreño

Received: 29-06-2018 Accepted: 26-04-2019

(2000), Escurra (2003) and Vázquez (2006) found general problems in the information reorganization and the inferential dimensions, while other authors obtained better results in all aspects (J. García, 2009; Mazurca, 2004; Pérez-Zorrilla, 2002), particularly in the literal dimension (Montanero, 2004; Saenz & Fuchs, 2002; Stein & Trabasso, 1981).

Textual Typology

The aforementioned reading processes depend on the type of text that is involved. Although the textual typology is considered one of the external factors that takes part in reading comprehension, there are texts in which the connections that the reader makes between the information extracted and his or her background knowledge are fundamental for their comprehension (expository texts) and others in which the structure of that textual information predominates (literary texts). For this reason, in this study different text types from different areas of the primary education curriculum have been addressed, being encompassed in four segments: on the one hand, literary and expository texts, and on the other, those of continuous and discontinuous formats.

Regarding the literary text, when the reader is presented with commonplace situations, this comes with a mental framework that precedes what is going to be read, with the prototypical components of that text, its situation or concept and the relationships that exist between them, which is why only an updating of this mental model is needed.

The textual structures are basically those that are placed in the forefront in the comprehension of literary texts, leaving background knowledge in second place. The opposite occurs in informative texts, where it is background knowledge that makes it possible to interpret them. It thus seems logical to think that literary texts, above all narrative texts, should generally be easier to understand (Borda, 2000).

Expository texts, meanwhile, present students with different aspects concerning many other subjects of interest. They are the basis for understanding certain motivations of the society in which the student is immersed. It could be said that in expository texts there is usually an introduction, a problem, a solution, an evaluation, and a conclusion, as well as a series of logical and hierarchical relations between themes and subthemes that demand attentive reading in order to be understood.

Knowing this structure helps in the comprehension of these texts, but having background information on the content being discussed is fundamental for comprehending better.

In addition, a basic distinction that is present in this study is established between continuous texts and discontinuous texts. The former are composed of paragraphs, inserted into other larger structures. They are classified above all by their rhetorical objective or text type (narration, exposition, description, argumentation, instruction, document or record and hypertext). Discontinuous texts, meanwhile, refer to graphics, tables, diagrams, and so on. Their information layout is very different to continuous texts, which is why their reading cannot be done linearly.

Reading Comprehension in Bilingual Students

The advantages that bilingual people have are indisputable at present: the bilingual person is capable of thinking with greater flexibility and creativity than their monolingual counterpart (Baker, 2014; O. García, 2009). Furthermore, the bilingual is aware of the existence of two language systems, which benefits them in understanding the relation between an oral system and a written one, an advantage for learning to read. However, this skill may be difficult to develop for the bilingual, depending on certain conditions and the languages they use.

The multidimensional condition of the reading process (involving cognitive, linguistic, textual, sociocultural and biological variables), combined with the different ways of approaching a text (depending on the reading dimensions and the characteristics of the text types mentioned above), means that it is not at all easy to determine the scope of reading comprehension in the area of teaching and learning of second languages (L2) related with this study.

This complexity makes it difficult for the reader to employ, in their L2, those reading strategies that they created when acquiring this skill in their L1. The mechanisms of textual cohesion and the formulae for the organization and distribution of the ideas in a text are natural in one language but can prove to be forced for a non-native reader accustomed to another type of arrangement, which is why reading competence can depend more on familiarity with the textual conventions than with the actual level of linguistic competence (Acquaroni, 2004; Bialystok, 2013),

Received: 29-06-2018 Accepted: 26-04-2019

although some authors, such as Hudson (2007) indicate that knowledge of the L2 plays a more important role in comprehension than the actual ability to read in L1. In fact, Koda (2008) states that the comprehension of texts in L1 and L2 is very different, whether due to the reading process or the reading product.

In relation to this, studies such as those carried out by Kato (1999), Kleiman (2004), Klett (2005), Salazar (2006) and Specht (2010) have stressed the importance of the strategies developed in the L1 for the learning of the L2, and vice versa, so that they are mutually beneficial.

However, Hudson (2007) and Koda (2008) mention two factors that make L2 reading comprehension difficult: the existence of conceptual gaps between the necessary background knowledge and the excessive dependence on the sociocultural knowledge of the L1, particularly when the culture of origin and the second culture are very different. Similarly, Parodi (2005) stresses the difficulty that non-natives may have with the inference of the textual meanings due to their cultural and ideological conditioning, independently of other social factors that intervene in the development of the bilingual's reading skill (Bialystok, 2013). The diglossic situation in Melilla may be such a factor: Spanish is the official language and the only one used in education through linguistic submersion (though not disregarding the culture of origin), while *Tamazight* is the language of communication in family and informal settings, giving rise to a subtractive bilingualism among the Berber population (Mohamedi-Amaruch, 2018).

These difficulties for the transfer between the L1 and L2 are increased in the case of the participants of this study, for whom Spanish is their L2. Not only do they belong to a very different culture, but also it is impossible for them to transfer strategies from their L1 (Berber) to reading in Spanish, as their L1 does not have written representation for its speakers. The *neo-Tifinagh* script is only known by scholars of the language (Rico-Martín, 2004). Other similar studies draw comparable conclusions, including one that is closely related, in which the participating subjects are also from the Maghreb and ignorant of their mother tongue's written code: "[Zhora] presenta grandes dificultades para descifrar y entender el lenguaje gráfico por no estar familiarizada con el mismo y extrapola [erróneamente] su propia experiencia para explicar situaciones de un mundo referencial no conocido por ella" (Aguirre, Villalba, Hernández, & Najt, 2008, p. 182).

Salazar (2006), in her study on interdependence and transfer of linguistic skills between an L1 and a foreign language, showed that skill transfer is based, among other things, on the linguistic interdependence hypothesis, and that the typology of transfers responds to classification parameters relating to utilitarian character and to task similarity, in this case those referring to the reading process. This situation comes alongside the difficulty mentioned above that this Berber group faces in the reading process, due to the lack of an underlying skill common to both languages for this specific transfer for bilinguals that was indicated by Cummins (1978).

Moreover, there are many studies that have compared the reading comprehension of bilinguals and monolinguals with very diverse results: with no notable difference between the two groups (Canales, Velarde, Meléndez, & Lingán, 2014; Martín-Pastor, 2012; Medina & Aparicio, 2014), or with a higher reading level for monolinguals (Godoy, 2013; Huguet, Navarro, & Janés, 2007). As stated at the start of this section, one must not forget the multidimensional condition of the reading process when it comes to explaining the divergence of these results.

The Present Study

The aim of this study is to contrast the reading comprehension level of monolingual students (Spanish) and bilingual students (Berber-Spanish), who are finishing Primary Education in the Autonomous City of Melilla, a place with unique characteristics due to the bilingualism with diglossia of one of the majority cultural groups.

The specific objectives of this study are:

1. To find out and contrast the reading comprehension level of the monolingual and bilingual groups.
2. To distinguish the levels of every reading dimension in both groups.
3. To check the comprehension level of the students in the different text types.

Attaining these objectives will help the educational community decide what lines to follow in order to work on the less developed dimensions and the more difficult text types for students, with the intention of improving not only reading comprehension but all communicative skills.

Received: 29-06-2018 Accepted: 26-04-2019

Method

Quantitative research has been carried out with an empirical-analytical methodology, and a descriptive, *ex-post-facto* study with a diagnostic purpose, a type of research that is widely used in the field of education.

Participants and their Contexts

To strengthen educational competences and to train the students so that they progress appropriately on to the secondary-school stage is the goal of Primary Education. For this reason, this study deals with students in the final year of this basic stage (Level 6, age equivalent to UK Year 7) from ten schools in Melilla. It uses an intentional non-probabilistic sampling from an invited sample of 1130 students, reduced to 612 (54.16% of the total population) having ruled out the students with special educational needs and those schools that declined the invitation.

Looking at the personal variables that this study is concerned with, the distribution of this sample according to cultural origin is of 255 students (41.66%) of peninsular or European origin, and 357 (58.33%) Berbers. In terms of first language, 419 students (68.46%) are Spanish monolingual, and 193 (31.53%) are bilingual with Berber-L1.

Melilla is a small Spanish city located in the north of Africa. It has a population of 86,384 inhabitants belonging to very different cultures, the majority being of European origin and Berber.

The city presents a diglossic situation, in which although everyone uses Spanish as the official language and as the language used in education, *Tamazight* is the mother tongue and family language of the Berber culture. This is a thousand-year-old language belonging to the Lybico-Berber branch. For its users it is an oral language, since they are ignorant of its script, *neo-Tifinagh* (Rico-Martín, 2004).

With the passage of time and the evolution of intergroup relations, a certain subtractive bilingualism has been observed in the city, wherein *Tamazight* is spoken on the streets, but among groups of Berber children interaction in Spanish is increasing at the expense of the mother tongue. For more information on the sociolinguistic context of the city and the educational situation of its schoolchildren, it is worth reading the study by Mohamedi-Amaruch (2018).

Data Collection Tool

The Avaluació de la Comprensió Lectora (ACL-6) [Reading Comprehension Assessment] by Català *et ál.* (2008) is a standardized test designed to assess reading comprehension using diverse texts belonging to the Primary Education curriculum and with answers to items that bring together the aforementioned four reading dimensions.

The reason for choosing this tool is the variety of texts and reading levels it employs, making it one of the most apposite tests for carrying out a diagnostic-type descriptive analysis of the reading comprehension of students.

The ACL-6 is a test that has been validated in different bilingual and monolingual schools in Catalonia and Madrid, and also subsequently employed in other studies outside of Spain (Lago, Moreno, & Domínguez, 2007; Sandoval, Frit, Maldonado, & Rodríguez-Alveal, 2010).

It consists of ten texts of varied typology (narrative, expository, poetic, and data interpretation or discontinuous texts), with 36 dichotomous items specific for Level 6. Two of them (ACL-6.7a-b and ACL-6.9) were tailored to the cultural context of the city, without affecting the reliability of the test.

Cronbach's alpha was calculated as $\alpha = .803$, which guaranteed the instrument's reliability. The reading comprehension assessment is carried out by direct scoring intervals or correct answers corresponding to the scales of one-to-ten whit this interpretation: from 1-2, *very low levels* (0-8 correct answers), to 10, *very high level* (31-36 correct answer) (Catalá *et ál.,* 2008), which ensures the objectivity and unity of criteria in the assessment.

Procedure

The teachers of each Year 6 group, together with the researchers, were responsible for completing the ACL-6 booklet after joint sessions to determine the modification of the mentioned texts and the way to apply the test. These were carried out in the first morning session on two consecutive days. Before this, on the first day, the procedure for the readings was explained to the students and they practised with the training texts.

Received: 29-06-2018 Accepted: 26-04-2019

In order to achieve the study objectives, the data analysis was undertaken with the *Statistical Package for Social Sciences* (SPSS), version 22.0. Firstly, descriptive statistics were used to describe the participant sample, the frequencies and the percentages with respect to the variables 'cultural origin' and 'first language', as well as to analyse the results obtained by the study participants in the ACL-6 test. Likewise, parametric inferential statistics are used, thus enabling bivariate analyses. For all comparisons, a *p*-value of ≤.05 was considered statistically significant.

As a first reference for the comparison of results, we have taken the global score obtained by the students who participated in the original test by Català *et ál*. (2008), and as the next main reference points, the global percentage of the correct answers obtained by the Melilla students and the mean percentage obtained by each group (monolingual and bilingual).

Results

Results of the ACL-6 Test

The descriptive statistical analysis of the ACL-6 test shows that the mean of the results obtained by all the participants was **13.68 points**, a low value considering that the scores could go from 0 to 36.

Thus it was discovered that 69.9% of participants were placed in low levels of reading comprehension (1-4 on the one-to-ten scale), 18.9% were in the middle values of comprehension (5-6 pts.) and only 11% of the students reached higher levels of reading (7-10 pts.). What is notable is that practically 70% of the city's Year 6 students attain a grade of reading comprehension below the mean expected for their year.

When forming comparisons according to the cultural origin of the students through Student's *t* test, very significant differences could be observed and with a large effect size, depending on whether a student belongs to one or other of the groups, with the students of Peninsular Spanish origin obtaining better scores in reading comprehension than those of Berber origin ($M_{Peninsular}$= 16.62; M_{Berber}= 11.58; *t*= 11.42; *d*= 0.94; *p*=.000; *gl*= 610).

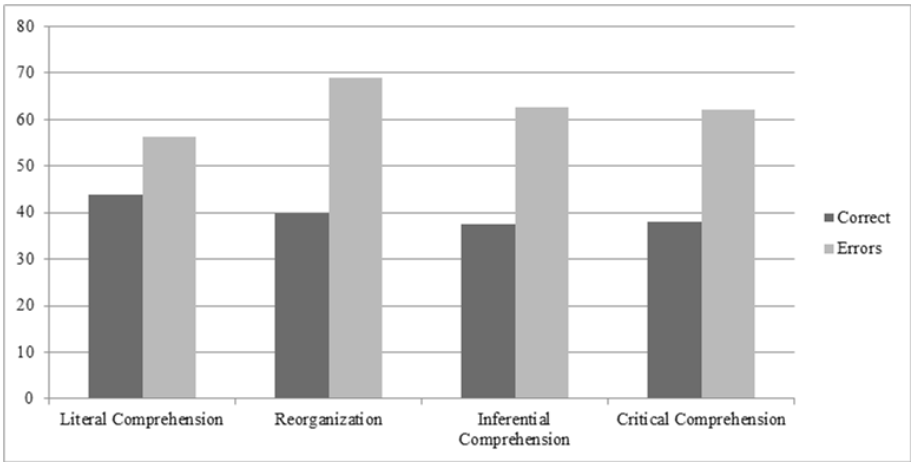
But the objective of this study does not depend so much on cultural origin as on the first language of the students. Therefore, proceeding to

study the existing relationship between the reading comprehension of the participating students and their first language, Student's *t*-statistic produced a very significant difference between both variables with a medium effect size, the highest level being the reading comprehension of the monolingual students ($M_{Spanish}$= 14.62; M_{Berber}= 11.64; *t*= 5.96; *d*= 0.52; *p*=.000; *gl*= 610).

Results of the ACL-6 Test as a Function of the Reading Comprehension Processes

The global positive results of the study with regard to the various comprehension levels did not reach the mean score, as observed in Graphic I in the difference between the students' correct answers and their errors.

GRAPHIC I. Global results of the test as a function of the comprehension processes (% of correct answers and errors)



Source: authors.

In all the types of reading comprehension, very significant statistical differences have been found, using Student's *t* test, between the mean

Received: 29-06-2018 Accepted: 26-04-2019

scores, with the monolingual students always obtaining the best results, but the differences have a small effect size in inferential and critical comprehension, and moderate or close to moderate in literal comprehension and in the processes of information reorganization (see Table II).

TABLE II. Mean scores in ACL-6 according to the reading comprehension processes of monolingual and bilingual students

Comprehension Processes	First Language				t	d
	Spanish (n= 419)		Berber (n= 193)			
	M	SD	M	SD		
Literal Comprehension	5.14	2.62	4.07	2.22	4.90*	0.43
Information Reorganization	2.71	1.53	1.94	1.26	6.08*	0.53
Inferential Comprehension	5.10	2.51	4.38	2.39	3.32*	0.29
Critical Comprehension	1.65	1.14	1.21	1.13	4.47*	0.39

Note: * $p \leq .001$; $df = 610$. Source: authors.

Results of the ACL-6 Test as a Function of the Textual Typology

In parallel with the analysis of the students' comprehension levels according to the ACL-6 test, the results were examined as a function of the type of text used and its different comprehension requirements. The scores achieved by the students show that the percentages of correct answers and errors are similar in the three text types, being practically equal in the informative and interpretative texts (see Table III).

TABLE III. Results of the ACL-6 test according to text type

Text Type	% correct	% errors
Literary texts Narrative & poetic	39.18	60.8
Informative texts Expository	37.13	62.86
Interpretive texts Graphics & data	37.15	62.85

Source: authors.

On analysing the differences of reading comprehension according to these texts and the monolingual or bilingual condition of the students, *t* again shows, in Table IV, very significant statistical differences between the mean scores, in which the monolingual group again stands out with the highest means in all the texts, although with a weak size effect.

TABLE IV. Mean scores in ACL-6 according to the different texts worked with

Texts worked	First Language				t	d
	Spanish (n = 419)		Berber (n = 193)			
	M	SD	M	SD		
Literary Texts	6.28	3.13	5.00	2.84	4.84*	0.42
Expository Texts	3.52	1.58	2.94	1.50	4.22*	0.36
Texts with Graphics & Data	4.81	2.74	3.67	2.43	4.95*	0.43

Note: * p = .000; df = 610. Source: authors.

Discussion

Reading Comprehension according to the ACL-6 Test

If the results obtained in Melilla (13.68 points) are contrasted with those from Català *et ál.* (2008) (19.2 points), a difference of almost six points

Received: 29-06-2018 Accepted: 26-04-2019

is observed. Furthermore, the data analysis shows that practically 70% of the students present a degree of comprehension below the average expected for their year. These data are similar to those obtained by Spain in PISA 2009 (IE, 2010a), the last assessment of the students of Melilla and in which the city occupied last place in the country.

With regards to the comparison of the ACL-6 results as a function of first language, the level of reading comprehension of the monolingual students in higher, with statistically significant differences but of medium effect: those who have *Tamazight*-L1 obtained a score of 11.64, three points lower than those obtained by their peers with Spanish-L1, with a score of 14.62 points.

In the introduction, it was stated that there are few studies on the linguistic competence of Spanish-Berber bilinguals. Those that exist, Herrera and Defior (2005), Herrera *et ál.* (2009), and Olmos (2009), have a participant sample of children of a different age to those in this study. They made evident the worse performance of these students compared to monolingual children in the main curricular subjects in general and in reading skills in particular. The problem lies not so much in the diglossic situation or in the linguistic submersion in Spanish of the education system, but in the socio-economic and cultural characteristics of this social group, the low level of education of their older members, and the weak family guidance that exists among them.

It is, furthermore, impossible at times to assess competences in their first language in order to establish a comparison of their mastery of the two languages that they use. In the cases of children who have a delayed start to education, they have not even been formally taught to read and write in their country of origin and have only attained productive and/or receptive oral competence, meaning that it is impossible to know whether their performance in their L1 is deficient or not because there is still no standardized test to measure it, since it is not a language that has been standardized.

Outside the sphere of the city, the results of the study are not in agreement with those from the study by Martín-Pastor (2012), which similarly compared reading comprehension between Spanish-speaking students and immigrants with a first language other than Spanish. In this case, the differences in the scores were not significant.

However, some studies, such as that by Huguet, Navarro and Janés (2007), have demonstrated that foreign students do not come to match

their Spanish counterparts in mastery of the language, despite their stay in Spain being longer than six years, the period in which it is understood that a person can obtain sufficient *cognitive academic language proficiency* (Cummins, 1979), necessary for learning tasks such as those of a linguistic type and reading comprehension. The study by Godoy (2013) draws a similar conclusion when examining the language and socialization of immigrant students as facilitators of the learning process, and shows that these students' scores are always lower than those of the Spanish students in all aspects of the language. Studies on a national and international level have similar results with regard to academic performance in Spanish Language (including reading comprehension) and Mathematics (IE, 2010a, 2010b).

However, authors such as Kato (1999), Kleiman (2004), Klett (2005) and Specht (2010), in their research on students whose first languages are different to the language of education, assert that knowledge of the linguistic structure and reading activity in their L1 aids the learning and comprehension of other languages, in the same way that the learning of an L2 enables a greater mastery of the L1 in that it requires intentional reflection about linguistic structures. Salazar (2006), meanwhile, advocated linguistic interdependence and task similarity to enable the transfer of linguistic strategies –reading strategies, in this case– during the learning of the L2. Nevertheless, the case of the Melilla bilingual students (L1-Berber and L2-Spanish) is not comparable with these situations due to the particular oral nature of the *Tamazight* language and its linguistic distance from Spanish. This means that these bilingual students do not have the underlying shared skill that Cummins (1978) had shown to enable the transfer of learning techniques between two languages, as this skill does not exist for the written code of Spanish for these Berber students. This is aside from the fact that there may be other familial and social determinants that are involved in these inter-group differences, as Bialystok (2013) clearly indicates, as well as Mohamedi-Amaruch (2018) in his analysis of the sociolinguistic and educational characteristics of Melilla.

Received: 29-06-2018 Accepted: 26-04-2019

The Reading Comprehension Processes

As a summary of the ACL-6 global results analysis, the difficulties that a large proportion of the students find when understanding a text literally are shown (56.26% errors in literal comprehension), being above all the most basic reading dimension. Even so, it is the dimension with the best results. Similarly, the students find it hard to establish logical connections between the text data in order to interpret its content (62.47% errors), just as they have problems in evaluating and giving their opinion on any text (62.1% errors), which leads to ensuring the lack of command of the reading skills involved in inferential and critical comprehension. This is along the same lines as the results shown by Kendeou *et ál.* (2014) for higher-level reading processes, such as making inferences. It is important to observe that inferential reading is the dimension that is given most attention in Spanish reading comprehension assessments.

But the least developed dimension in the Year 6 group is the reorganization of information (69% errors), which hinders the interpretation of any text read, above all with graphics, tables and data, although they also had difficulty in classifying the information from the literary texts, leading them to passive rather than active reading.

Another study, Vázquez (2006), obtains similar results when applying the same ACL-6 test with a group of Year 6 Mexican students and, moreover, shows a null score for these students in the interpretative and critical reading dimensions.

The results of the Melilla study also agree with those obtained by Carreño (2000) and Escurra (2003) with students of the same level. In contrast, they differ from those found by Pérez-Zorrilla (2002), also obtained in a similar subtractive bilingual context. This author indicates better achievements, although still insufficient, in the information reorganization ability (45% incorrect answers), followed by inferential comprehension (47%), critical (52%) and lastly the literal dimension (59%). All of these studies are in contrast to those by Marzuca (2004) and J. García (2009), which both found very positive results in all reading dimensions.

The results of this work reveal that there are differences between the two participant groups in all reading dimensions, with the monolingual group always attaining better results, despite their low reading levels. Although a similar degree of significance exists in the contrast of means

in all the dimensions, it is noteworthy that these differences are more pronounced, having a medium effect in literal comprehension and in information reorganization and less in critical and inferential reading, levels with significant statistical differences but of weak effect.

Those students whose first language was Berber have greater difficulty in a reading type that demands greater depth and skill of thought –for example, when faced with the decision of which of several possibilities states the purpose of a text, they tend to choose those options that refer to purpose in general, without considering the context of the text.

They also have problems ranking the information of a text, and they fail in tasks that require them to take into account information that is distributed throughout the text because they lack the sufficient capacity to make inferences in the reading. Kendeou *et ál.* (2014) show the difficulty presented by this type of reading comprehension. Furthermore, Parodi (2005) and Aguirre *et ál.* (2008) proved that the cultural and ideological conditioning of the reader could hinder inference in reading in a language other than the first. This may be the case with students of Berber origin, whose culture is very different to that reflected in the texts they work with –hence some are adapted– even though these students, given their academic level and their contact with other cultures in the city, should not have so many difficulties in these aspects.

Likewise, they struggle with the task of searching the text for information that fulfils various requirements or to connect their background knowledge to interpret the textual information. They therefore also have problems making critical judgements on the text content using their knowledge of life. These difficulties affect the interpretation of the textual typology they are working on.

Outside of Spain, also in diglossic contexts, studies can be found that show significant differences in pre-reading and psycholinguistic abilities between primary-school students who are monolingual in Spanish and incipient bilingual in Quechua and Spanish, in favour of the former. However, there were no such differences when it came to the actual reading processes (Canales *et ál.*, 2014). In another Peruvian diglossic context, last-year primary students did not show significant differences in reading processes between advanced bilinguals and monolinguals, although they did in the rates of accuracy and reading speed (Medina & Aparicio, 2014).

The Relationship between Text Types and Reading Comprehension

After the analysis according to the textual modality employed, it is established how there is hardly any difference between the comprehension of the literary, expository and interpretative texts for the participants as a whole. In the last two text types the results were slightly worse. Nevertheless, the global results are not what one would wish for in Year 6 students, only 30% of these students understand the explicit information of the texts, though in their most literal and explicit aspects, which do not require complex background knowledge, as can be the case for expository texts. Along the same lines as these global results are the works of Stein and Trabasso (1981), Saenz and Fuchs (2002), and Montanero (2004).

As has already been mentioned, the comprehension of an expository text presents more difficulties than the narrative (literary) text (Borda, 2000). As can be seen in the results, the bilingual group give even fewer correct answers than the rest of the students in all types of texts, although in the literary texts their comprehension improves without attaining a score desirable for their academic level. The highest differences in the means of the groups are in the literary texts and the lowest are in the expository texts. Nevertheless, they are still of weak effect.

The results mentioned lead to observing significant differences in favour of the monolingual group in the factors analysed, linked to reading comprehension (comprehension processes and textual typology). However, it is true that, just as has been indicated by González (1983) and Medina and Aparicio (2014) in their studies, these differences could be explained by the fact of including incipient bilingual students, who present greater interferences of the L1 on the L2. Moreover, in the case of this study, these interferences occur at different language levels that can affect these students' reading comprehension (Rico-Martín, 2004).

Conclusion

In this study we have found that reading comprehension one cannot transfer from the first language in all cases. The elements of textuality are unique to each language and it is impossible to extrapolate them if the interlinguistic divergences are too great.

The analysis carried out has produced empirical data on the reading comprehension level of the Year 6 bilingual and monolingual students in Melilla. In relation to the test results and the objectives set out at the beginning, we conclude the following:

1st objective, *to find out and contrast the reading comprehension level of the monolingual and bilingual groups*. The performance of the Year 6 students in general is below what is expected for their stage in 69.9% of students, but the degree of comprehension is lower for the Berber-Spanish bilinguals, the differences being very significant according to group.

2nd objective, *to distinguish the levels of every reading dimension in both groups*. According to the global results, more than half the students have problems in the literal dimension, but they are worse in the inferential and critical dimensions. If this is worrying, even more so is the degree of comprehension that enables the reorganization of read text, as the inability of 69% of students in this skill was detected. When comparing both groups (monolingual and bilingual) in all the reading dimensions, the monolingual Spanish students perform better than the Berber-Spanish bilinguals, the differences being very significant, although in some cases (literal comprehension and information reorganization) they are of moderate and in others (inferential and critical comprehension) weak effect size.

3rd objective, *to check the comprehension level of the students in the different text types*. Although we confirm that there are no significant differences in the low comprehension of the students in general in the typology that was worked with, when done by groups, the monolingual students continue to perform better in the three text types analysed: literary, expository and discontinuous (graphics and data tables).

Related to these differences between the groups, and the disadvantageous position of the bilingual, it must be noted that in this geographical setting of Melilla, a significant sector of the population needs two languages for daily life. In usual situations, having a stimulating input regarding the different communicative skills in both languages would be highly beneficial; but this is very difficult when they are so divergent that the L1 has no graphic or written representation known to the students

and, even if there were, the writing system is very different to the Spanish they learn in the classroom. With a situation of submersion in the Spanish language and as the Berber population acquires more education, the use of their L1 becomes more and more restricted to the family environment, meaning that in many cases this is subtractive bilingualism.

Moreover, when there are social situations that mean that this bilingualism is not developed as it should be, it makes an action plan necessary in the form of psycho-educational intervention programmes based on the bilingual condition to work on the cognitive development and Spanish-L2 learning processes of these students, as the language of teaching.

To this end, the Spanish Ministry of Education in Melilla, concerned about the city's excessive school-dropout rates, is conducting a series of programmes (Linguistic Immersion Programme, Educational Support Programme and School Support Programme for districts IV and V[1]). Mohamedi-Amaruch (2018) details all of these actions, both current and those undertaken previously. The ALEHOP project (Molina, 2015) must also be mentioned, which works with the youngest students on oral expression, the basic skill before tackling literacy skills, along with reading comprehension in children with a non-Spanish L1. All of this is added to a programme of remedial education that is carried out in all schools to develop the linguistic communicative competences of students with a low level of Spanish. Thus it is hoped to eliminate the disadvantages of this bilingual group.

References

Acquaroni, R. (2004). La comprensión lectora. In J. Sánchez, & I. Santos (Ed.), *Vademécum para la formación de profesores. Enseñar español como segunda lengua (L2)/lengua extranjera (LE)* (pp. 943-967). Madrid: SGEL.

Aguirre, C., Villalba, F., Hernández, M.T., & Najt, M. (2008). *Estudio comparativo entre la adquisición del español como primera lengua y*

(1) These districts, with a population of Berber origin, are the most disadvantaged of the city.

la adquisición del español como segunda lengua para su aplicación metodológica en la enseñanza del español a inmigrantes (Dosieres Segunda lengua e Inmigración, 1). Retrieved from http://www.segundaslenguaseinmigracion.es/

Baker, C. (2014). *A Parents' and Teachers' Guide to Bilingualism*. Clevedon, UK: Multilingual Matters.

Barrett, T.C. (1968). Taxonomy of Cognitive and Affective Dimensions of Reading Comprehension. In H.M. Robinson (Ed.), *Innovation and Change in Reading Instruction* (pp. 17-23). Chicago: University of Chicago Press.

Bialystok, E. (2013). The Impact of Bilingualism on Language and Literacy Development. In T.K. Bhatia, & W.C. Ritchie (Ed.), *The Handbook of Bilingualism and Multilingualism* (pp. 624-648). Oxford, UK: Blackwell.

Borda, M.I. (2000). Estrategias de comprensión lectora en los talleres de lectura de los cuentos infantiles. *Lenguaje y textos, 15*, 51-59.

Canales, R., Velarde, E., Meléndez, M., & Lingán, S. (2014). Diferencias en habilidades psicolingüísticas y lectura en niños bilingües quechua-castellano y monolingües castellano de primer grado. *Theorema, 1*(1), 99-110.

Carreño, M. (Ed.) (2000). *Teorías e instituciones contemporáneas de educación*. Madrid: Síntesis.

Català, G., Català, M., Molina, E., & Monclús, R. (2008). *Evaluación de la comprensión lectora*. Barcelona: Graó.

Consejo Escolar del Estado [State School Council] (2016). *Informe 2016 sobre el estado del sistema educativo*. Madrid: Ministerio de Educación, Cultura y Deporte.

Cummins, J. (1978). Educational Implications of Mother Tongue Maintenance in Minority-language Groups. *The Canadian Modern Language Review, 34,* 395-416.

Cummins, J. (1979). Cognitive/Academic Language Proficiency, Linguistic Interdependence, the Optimum Age Question and some other Matters. *Working Papers on Bilingualism, 19*, 197-205.

Escurra, M. (2003). Comprensión de lectura y velocidad en alumnos de sexto grado de primaria de centros educativos estatales y no estatales de Lima. *Persona, 6*, 99-134. Retrieved from http://www.redalyc.org/articulo.oa?id=147118110006

Eurydice (2011). *Teaching Reading in Europe: Context, Policies and Practices*. Brussels: Eurydice-European Commission.

García, J. (2009). *Importancia de la comprensión de textos en los estudiantes* (Doctoral Dissertation). University of Santiago de Compostela, Spain.

García, O. (2009). Bilingual Education in the 21st Century: A Global Perspective. Malden, MA: Wiley-Blackwell.

Godoy, M.J. (2013). *El lenguaje y la socialización del alumnado inmigrante como facilitadores del proceso de aprendizaje. Estudio psicométrico y curricular* (Doctoral Dissertation). University of Extremadura, Spain. Retrieved from http://hdl.handle.net/10662/376

Goldman, S.R., McCarthy, K.S., & Burkett, C. (2015). Interpretive Inferences in Literature. In E.J. O'Brien, A.E. Cook, & R.F. Lorch (Ed.), *Inferences during Reading* (pp. 386-415). Cambridge, UK: Cambridge University Press.

González, R. (1983). Problemas psicológicos en la comunicación lingüística en el Perú. *Revista de Psicología, 1*(1), 53-60.

Herrera, L., & Defior, S. (2005). Procesamiento fonológico en niños con lengua materna española y *tamazight*. In J.M. Oros, J. Anderson, & J. Varela (Ed.), *La enseñanza de las lenguas en una Europa multicultural* (pp. 311-322). Santiago de Compostela: University of Santiago de Compostela.

Herrera, L., Defior, S., Serrano, F., & Jiménez-Fernández, G. (2009). Influencia de la procedencia cultural (europea vs. amazight) en las habilidades de procesamiento fonológico, de vocabulario y de lectura de letras. *Publicaciones, Facultad de Educación y Humanidades de Melilla, 39*, 43-60.

Hudson, T. (2007). *Teaching Second Language Reading*. Oxford, UK: Oxford University Press.

Huguet, Á., Navarro, J.L., & Janés. J. (2007). La adquisición del castellano por los escolares inmigrantes. El papel del tiempo de estancia y la lengua familiar. *Anuario de Psicología, 38*(3), 357-375.

IE (Instituto de Evaluación [Institute of Evaluation]) (2010a). *PISA 2009. Programa para la evaluación internacional de los alumnos de la OCDE* (Informe español). Madrid: Ministerio de Educación.

IE (Instituto de Evaluación [Institute of Evaluation]) (2010b). *Evaluación general de diagnóstico 2009. Educación Primaria. Cuarto curso* (Informe de resultados). Madrid: Ministerio de Educación.

INEE (Instituto Nacional de Evaluación Educativa [National Institute of Educational Assessment]) (2012). *Ítems liberados. Evaluación de Educación Primaria. Versión 1.0*. Madrid: Ministerio de Educación, Cultura y Deporte. Retrieved from https://www.mecd.gob.es/inee/dam/jcr:3e26c9ab-8084-4d34-99dd-890f726f55b2/itemsliberadoseducacionprimaria.pdf

Kato, M. (1999). *O aprendizado da leitura. Texto e Linguagem*. São Paulo: Martins Fontes.

Kendeou, P., van den Broek, P., Helder, A., & Karlsson, J. (2014). A Cognitive View of Reading Comprehension: Implications for Reading Difficulties. *Learning Disabilities Research & Practices, 29*(1), 10-16.

Kintsch, W. (2003). *Comprehension: a Paradigm for Cognition*. Cambridge, UK: Cambridge University Press.

Kleiman, A. (2004). *Texto e leitor: aspectos cognitivos da leitura*. Campinas. São Paulo: Pontes.

Klett, E. (2005). La relación lengua materna/lengua extranjera: aspectos didácticos y psicolingüísticos. In E. Klett (Ed.), *Didáctica de las Lenguas Extranjeras: Una agenda actual* (pp. 51-69). Buenos Aires: Araucaria Editora.

Koda, K. (2008). Impacts of Prior Literacy Experience on Second-Language Learning to Read. In K. Koda, & A.M. Zehler (Ed.), *Learning to Read across Languages: Cross-Linguistic Relationships in First- and Second-Language Literacy Development* (pp. 68-96). New York: Routledge.

Lago, C., Moreno, J., & Domínguez, M. (2007). *Adaptación al contexto chiapaneco-indígena del manual de la evaluación de la comprensión lectora.* Pruebas ACL (1.º-6.º *de Primaria)*. San Cristóbal de las Casas, México. Retrieved from http://filosofiaparaninos.org/prueba-de-comprension-lectora-en-chiapas/

Martín-Pastor, E. (2012). *Inclusión del alumnado inmigrante: evaluación de su comprensión lectora* (Doctoral Dissertation). University of Salamanca, Spain. Retrieved from http://hdl.handle.net/10366/121450

Marzuca, R. (2004). *El Programa de Lectura Silenciosa Sostenida y su efecto sobre la comprensión lectora* (Doctoral Dissertation). University of Chile, Chile. Retrieved from http://www.repositorio.uchile.cl/handle/2250/105945

Medina, G.J., & Aparicio, P.M. (2014). Procesamiento lector en niños bilingües avanzados quechua-castellano y monolingües castellano circunscrito en dos provincias del Departamento de Ayacucho

(Huamanga y Vilcashuamán) (Master Dissertation). Pontifical Catholic University of Perú, Perú. Retrieved from http://tesis.pucp.edu.pe/repositorio/handle/123456789/5763

Mohamedi-Amaruch, A. (2018). *La comprensión lectora en alumnos bilingües mazigio-español al término de la Educación Primaria*. Melilla: UNESCO-C.A.M.

Molina, E. (2015). La mejora de la competencia en comunicación lingüística: "ALEHOP" y la experiencia en un contexto desfavorecido. *Participación educativa, 4*(6), 91-95.

Montanero, M. (2004). Cómo evaluar la comprensión lectora: alternativas y limitaciones. *Revista de Educación, 335*, 415-428.

Olmos, M.C. (2009). Aprender juntos entre alumnado con lenguas maternas de origen diferente pero de igual nacionalidad. *Revista Educativa Inclusiva, 2*(2), 41-50.

Parodi, G. (2005). *Comprensión de textos escritos*. Buenos Aires: Eudeba.

Pérez-Zorrilla, M.J. (2002). *Evaluación de la Educación primaria: sexto curso. Centros de Ceuta y Melilla*. Madrid: Ministerio de Educación, Cultura y Deporte.

Rico-Martín, A.M. (2004). Las hipótesis del análisis contrastivo y de la identidad referidas a errores en la adquisición del español-L2 de alumnos de origen mazigio. *Publicaciones. Facultad de Educación y Humanidades, 34*, 177-194.

Saenz, L.M., & Fuchs, L. (2002). Examining the Reading Difficulty of Secondary Students with Learning Disabilities: Expository versus Narrative Text. *Remedial and Special Education, 23*(1), 31-41.

Salazar, L. (2006). Interdependencia lingüística, transferencia y enseñanza-aprendizaje de lenguas extranjeras. *Laurus, 12*, 45-72. Retrieved from http://www.redalyc.org/articulo.oa?id=76109904

Sandoval, P.R., Frit, M., Maldonado, A.C., & Rodríguez-Alveal, F. (2010). Evaluación de habilidades en matemática y comprensión lectora en estudiantes que ingresan a pedagogía en educación básica: un estudio comparativo en dos universidades del Consejo de Rectores. *Educar em Revista, special number 2*, 73-102.

Solé, I. (2012). Competencia lectora y aprendizaje. *Revista Iberoamericana de Educación, 59*, 43-61. Retrieved from http://rieoei.org/rie_contenedor.php?numero=rie59

Specht, L.B. (2010). *Comprensión de lectura de texto en español por aprendices brasileños* (Master Dissertation). Federal University of

Paraná, Brasil. Retrieved from http://acervodigital.ufpr.br/bitstream/handle/1884/25884/Disertacao%20LUISA%20BARBARA%20SPECHT.pdf?sequence=1

Stein, N.L., & Trabasso, T. (1981). What's in a Story: An Approach to Comprehension and Instruction. In R. Glaser (Ed.), *Advances in the Psychology of Instruction* (Vol. 2, pp. 213-267). Hillsdale, NJ: Erlbaum.

van den Broek, P., & Helder, A. (2017). Cognitive Processes in Discourse Comprehension: Passive Processes, Reader-Initiated Processes, and Evolving Mental Representations. *Discourse Processes, 54*(5-6), 360-372. doi: 10.1080/0163853X.2017.1306677

Vázquez, M.E. (2006). *El nivel de comprensión lectora en sexto grado de Primaria* (Master Dissertation). National Pedagogic University, México. Retrieved from http://digitalacademico.ajusco.upn.mx:8080/tesis/handle/123456789/5899

Dirección de contacto: Ana M. Rico-Martín, Universidad de Granada, Facultad de Ciencias de la Educación y del Deporte de Melilla, Departamento de Didáctica de la Lengua y la Literatura. E-mail: amrico@ugr.es

Received: 29-06-2018 Accepted: 26-04-2019

Teacher's Quality in Colombia, Does Monetary Incentives Matter?[1]

Calidad Docente en Colombia ¿Una cuestión de incentivos económicos?

DOI: 10.4438/1988-592X-RE-2019-385-424

Ana Bolena Escobar Escobar
UNbralesa Brasilia
Jorge Leonardo Duarte Rodríguez
Instituto Colombiano de Bienestar Familiar
Sebastián Estrada Jaramillo
Secretaría de Desarrollo Económico, Alcaldía de Cali
Silvana Godoy Mateus
Instituto Colombiano de Bienestar Familiar
Ximena Dueñas Herrera
Banco Interamericano de Desarrollo

Abstract

There is a broad academic debate about the relevance of economic incentives designed to increase the quality of school education. Although research analyzing the relationship between incentive and performance shows mixed results, the implementation of this type of strategy has become popular around the world. In this context, the objective of this article is to estimate the impact of the Colombian program of economic incentives, which assigns bonuses to administrative and teaching staff of public establishments that achieve a certain threshold in the Synthetic Index of Education Quality (ISCE). In this article, the

(1) This work was developed during the job vinculation of some of the authors with the ICFES (Colombian Institute for the Evaluation of Education) between the year 2017 and 2018.

Received: 18-09-2018 Accepted: 26-04-2019

program design is used to construct two identification strategies that approximate *discontinuous regression* and *propensity score matching* methods. Under certain assumptions, the implementation of these strategies shows that for the subsample of schools tested the program does not have a positive impact on the change in the scores of the Saber 3, 5 and 9 test. In fact, one year after the implementation, the non-beneficiaries improved more than the beneficiaries, and two years later no significant differences were found between the two groups. This article also discusses the theoretical approaches that guide the understanding of the results obtained and invites to deepen the research on how the programs should be designed and implemented in order to increase teaching effort and academic performance.

Key words: natural experiment, incentives, school performance, standardized test, impact evaluation.

Resumen

Existe un amplio debate académico sobre la pertinencia de los incentivos económicos diseñados para aumentar la calidad de la educación escolar. A pesar de que las investigaciones que analizan la relación entre el incentivo y el desempeño evidencian resultados mixtos, la implementación de este tipo de estrategia se ha popularizado alrededor del mundo. En este contexto, el objetivo de este artículo es estimar el impacto del programa colombiano de incentivos económicos, que asigna bonificaciones al personal administrativo y docente de los establecimientos públicos que logren alcanzar cierto umbral en el Índice Sintético de Calidad Educativa (ISCE). En este artículo, se utiliza el diseño del programa para construir dos estrategias de identificación que se aproximan a los métodos de *regresión discontinua* y *propensity score matching*. Bajo ciertos supuestos, la implementación de estas estrategias muestra que para la submuestra de colegios evaluados el programa no tiene un impacto positivo sobre el cambio en los puntajes de la prueba Saber 3°, 5° y 9°. De hecho, un año después de la implementación, los no beneficiarios mejoraron más que los beneficiarios, y dos años después no se encuentran diferencias significativas entre ambos grupos. En este artículo también se discuten los planteamientos teóricos que orientan el entendimiento de los resultados obtenidos y se invita a profundizar en la investigación sobre la forma en la que deben diseñarse e implementarse los programas para lograr aumentar el esfuerzo docente y el desempeño académico.

Palabras clave: experimento natural, incentivos, desempeño escolar, pruebas estandarizadas, evaluación de impacto.

Introduction

A vast majority of research that analyzes the factors associated to the learning process, acknowledges teachers as one of the main components in explaining the level of achievement in student performance (Hanushek & Woessmann, 2011). Based on this approach, and on the recommendations of the OCDE[2], Colombia has supported the idea that improving the teacher´s remuneration scheme could enhance the academic performance of students. As a result, the Ministery of National Education (MEN) set up a program of economic incentives for teachers and administrative personnel in the public schools of the country. The beneficiaries of this particular program were defined in response to their goal fulfillment of the Minimum Annual Improvement (MMA)[3], stablished in accordance with the Synthetical Index of Educational Quality (ISCE)[4].

In this context, this article´s objective is to estimate the effect of the economic incentives program on the observed changes from one year to the next in Saber 3°, 5° y 9° test scores. In this case, theoretical elements are used to discuss the unexpected effects that the program may present. Based on the defined rules for the beneficiary's selection, two quantitative strategies are explored that approach two different methods: *discontinuity regression and score matching method*. Under certain assumptions, the implementation of these strategies allows us to select sub samples of beneficiary and non-beneficiary schools with similar characteristics, in which the differences between means would approximate the program´s effect, since it is possible to argue that the only difference between these groups relies solely on the incentive. The implementation of these strategies led us to conclude that on the sub samples of the tested schools, the program had no positive impact over the score changes in Saber test 3°, 5° y 9°.

Analyzing the use of economic incentives in the education sector from a theoretical and quantitative perspective, is of high relevance for the formulation of education policies (Shifrer et al., 2017), given the fact that it forms an idea about the possible consequences of this type of programs, and also allows to identify the characteristics that must be taken into ac-

(2) Organization for Economic Co-operation and Development.
(3) In Spanish: Mínima de Mejoramiento Anual.
(4) In Spanish: Índice Sintético de Calidad Educativa.

count in the formulation and implementation of the policy. Consequently, assessing the impact of the Colombian incentives program on the ground of academic performance, is an exercise that contributes to the debate about the relevance of this type of programs.

This article is organized in the following way: after this introduction, indicators that determine the beneficiaries of the program and how they are delivered are explained. Then, we present a theoretical and empirical background concerning the economic incentives program in the education sector. Later, we described the methodological strategies used and the results we found. At last, we expose a discussion and some general conclusions.

Description of the program

Indicators through which the beneficiaries are defined: Synthetic Index for Education Quality (ISCE) and Minimum Annual Improvement (MMA)

To measure the quality of school education, the MEN requested the Colombian Institute for the Evaluation of the Education (ICFES)[5], to create an index that could gather different results and was easily interpretable. As a result, the ICFES built the ISCE (Synthetic Index for Educational Quality). This indicator is generated within schools and cycles of education (primary, middle and high school), and is expressed on a 1 to 10 scale formed by four main components: Performance, Progress, Efficiency, and School Environment.

In each educational cycle, the score obtained by the schools in the two first components, is calculated based on the achieved results by the corresponding grades (for example, en primary school the grades that are considered are third and fifth grade) in the areas of math and languages in the Saber tests 3°, 5° y 9° (from now on Saber) and Saber 11. The first component has a 40% value on the ISCE and corresponds to the rescaled average of the scores obtained during a particular year in the two mentioned areas. The second component also corresponds to a 40% value in the ISCE but is obtained through the percentage change in the

(5) In Spanish: Instituto Colombiano para la Evaluación de la Educación.

Received: 18-09-2018 Accepted: 26-04-2019

percentage of students located in two of the four levels of performance in the Saber test: Insufficient and Advanced[6]. The School Environment component equals a 10% of the ISCE and corresponds to the value of two indicators that measure how appropriate is the classroom´s environment in terms of discipline, and how convenient are the feedback activities carried out in the classroom. The last component equals the total average of the approval rate in all the school grades that form the tested school cycle. This last component seeks to mitigate the perverse incentive of failing students with a low performance, as a way to prevent them from presenting the Saber test. It is important to clarify, that the ISCE published in a particular year, utilizes the data of Saber applied in the two previous years (Icfes, 2016).

The Ministry of National Education (MEN) as part of the national policy "*Colombia, la mejor educada de América Latina en el 2025*" (Colombia the most educated country in Latin America in year 2025), also requested to ICFES the creation of Minimum Annual Improvement (MMA) goals. This indicator was defined so that between 2015 and 2025, Colombia´s score in PISA test will reach the one obtained by Chile. To develop this task, a comparison between the scores of Saber and PISA during 2012 was made, considering that at the beginning of 2015 it was the last year with published results. Afterwards, a calculation was made on how much the national ISCE should increase so that in 2025 the average in Saber, expressed on the PISA scale, equals Chile´s average. After defining the national goal of 2025, the annual goals for schools were assigned through a logistic growth function that indicates how much the ISCE of each school should increase to reach the national objective (Icfes, 2017).

Eligibility and beneficiaries' selection

The duration of the program was two years, though it may seem longer due to the difference between dates in which; the program was announced, the beneficiaries were defined, and the incentive was delivered. In July 2015, few months after the first ISCE was published, the incentives program was launched, announcing that all teachers and administrative

(6) In the particular case of the test Saber 11, the first and last quintile of the positions are used, because this Test didn´t have performance levels when ISCE was designed.

personnel of the schools that reached or exceed the MMA would receive a price that would go from half a monthly salary to a full additional salary (MEN, 2015). At the beginning of 2016, the beneficiaries were announced, and until the middle of the year the economic incentives were delivered. During the first year of the program, the eligible schools were defined as official establishments ascribed to the *Programa Todos a Aprender* (PTA program), or to *Jornada Única* (JU)[7], while in the second year all official schools were defined as eligible for the program.

The rules that define the beneficiary status depend on the education cycles offered by each school. An establishment only becomes a beneficiary if it reaches or exceeds the MMA in all the cycles it offers. For example, a school that has the three educational cycles, and has only reached the MMA in primary, does not become a beneficiary, but one that only offers the primary education cycle, does become a beneficiary if it reaches the MMA of this cycle. This rule help to mitigate the perverse incentives of focusing only in the education cycles with the best chances of improvement.

In its two versions (2015 and 2016), the program delivered incentives ranging from half a month salary to an additional full month salary per year. Nevertheless, it is important to consider that the incentive is not the same for all, due to the variety of range in salary gain of the teachers based on the promotion ladder. The personnel of the beneficiary establishments earns half a salary if the school equals its MMA, and a full month salary if it exceeds twice the MMA which is defined as Goal of Excellence (MDE)[8]. Besides, if the school overcomes the MMA, but does not achieve the MDE, the school personnel obtains a percentage of the salary that is proportional to the achievement obtained.

(7) Jornada Única is a modality that stablishes a longer school day for students. The main objectives of Jornada Única are explained in the National Plan for Development 2014-2018 (Plan Nacional de Desarrollo 2014-2018) and in the Act 501 of 2016 (Decreto 501 de 2016). On the other side, the Program Todos a Aprender is a Colombian government initiative, that begins in 2011, whose goal is to strengthen teaching practice in low-performing education institutions, through a strategy of local accompaniment.

(8) In Spanish: Meta de Excelencia.

Received: 18-09-2018 Accepted: 26-04-2019

Academic Debate about incentives in education

Theoretical Approaches

Economy has studied incentives programs based on the called theory of agency. This approach states that: A principal (Ministry) engages an agent (teacher) to perform some service or action on their behalf (Jensen & Meckling, 1976). In this relationship between the principal and the agent, their could be uncertainty regarding the performance of the desired action due to a conflict of interests. In this scenario, the principal can limit divergences by introducing an incentive that aims to align the interests of both actors. Naturally, the implementation of this type of programs comes with some restrictions. Lavy (2009), highlights the importance of considering the possibility that all agents fulfil their goals because the incentives would represent a high cost to the organization. This author, also exposes the risk concerning agents leaving aside important activities that are not measured, and the difficulties associated with *free rider problem* (Holmstrom & Milgrom, 1991) when the incentive is assigned in a group. In reference to this last topic, is important to state that the alternative of delivering the incentive individually is disadvantageous when the desired action requires joint actions (Fryer, 2013; Goodman & Turner, 2013; Marsh et al., 2011).

According to Ryan (1982), in psychology, incentives are understood as rewards handed for achieving a performance goal set by a previous standard. This author also points out that the behavioral changes may also be explained through the intrinsic and extrinsic motivation; two concepts that reveal the complex and not necessarily positive relation between incentives and performance, because the performance can be strongly affected by the decrease in the level of intrinsic motivation. This phenomenon, known as the *crowding out* effect, happens when the incentive introduces two types of effects that can be antagonistic: the price effect rises extrinsic motivation because the behavioral change is mediated by a reward, and in contrast, the psychological effect can be detrimental to intrinsic motivation, due to the existence of an incentive that assumes that the task to be performed is not appealing or very expensive in terms of effort and capacity (Gneezy et al.,2011).

Empirical Evidence

Amongst the authors that have developed experimental studies that estimate the impact of group incentives in developing countries, we find Glewwe et al. (2010) in Kenia. These authors found a positive but not significant relationship between the incentive and performance of schools in standardized tests in this country. Another important study is the one from Muralidharan & Sundararaman (2011), who randomly assigned group and individual incentives to different schools in India. They found out during the first year, that group and individual incentives seem equally effective because both increase performance in language and math in a standard deviation of 0.12 and 0.16 respectively. Nevertheless, during the second year, individual incentives showed greater impact. For rural areas in Rajasthan, Duflo et al. (2012) built a structural model to evaluate the impact of an incentives program given to teachers of the non-official sector. In this case, the incentive given to promote teacher´s assistance to the establishment, increased the scores of the students in 0.17 standard deviation. In the same line of research but in the context of a developed country, Fryer (2013) analyzes an experiment in which an incentive was delivered randomly to 200 school in New York. These schools have the freedom to choose if the incentive was handed individually or to a group amongst their teachers. Fryer found out that the majority of schools opted for group assignment, and that, in average, the effect was significant and negative in secondary cycle, and non-significant in the elemental cycle. Although the results of all of these authors tend to differ from each other, there is a certain consensus concerning the importance of considering the free rider problem; these is concluded after the observation of major effectiveness of group incentives in small schools.

Regarding non-experimental studies, it is important to highlight Lavy´s (2002) work in Israel. This author estimates de impact of a monetary incentives program assigned to a group throughout a contest format, in which the magnitude of the incentive depends on the position of the participants. Lavy, takes advantage from the program´s eligibility criteria to apply the regression discontinuity method, from which he finds positive and significant impacts on students' performance in the Bagrut test [9]. This program was compared to an intervention developed simul-

(9) National Standardized Test to access superior education.

Received: 18-09-2018 Accepted: 26-04-2019

taneously in a sample of different schools, in which economic resources were delivered to extend the time of teaching and to offer training courses. The second program also showed a significant and positive impact, but in terms of costs it was less efficient. In this format of research, there are other evaluations made to program ASPIRE, implemented in Houston as a contest to deliver economic incentives at the level of school grades and subjects (for example fifth grade language) based on the added value of the score obtained in standardized tests. Imberman & Lovenheim (2015) use the design of this program to analyze if the effect of the incentive varies according to the number of students a teacher attends, as a total proportion of the total number of students of a same grade enrolled under a same subject**.** This last variable is understood as an intensity proxy of the incentive, because the bigger the number of students the teacher teaches, the bigger is the effect of their behavior over the probability of winning. Throughout the method of differences in differences (DD) and diverse controls, the authors concluded that the performance of students augments in response to the proportion of students that each teacher attends to, and, that the incentive effect, regardless of the intensity, is bigger among students with lower performance levels. Finally, the results of this particular study present evidence that favors designs in which the number of teachers from which the incentive depends is low. Because, in this way the number of students on which each teacher has responsibility increases.

Impact Evaluation

In this article, two main strategies are used that take advantage of the program´s design to estimate the impact of the incentive over the changes occurred in two consecutive years in Saber test scores. The first strategy is based on the *regression discontinuity method* (Lee & Lemieux, 2010), which consists on showing how different the variations in the scores are, between beneficiaries and non beneficiaries that were very close to the cut defined for receiving the incentive; in other words, the variations in scores between those schools with an ISCE close to MMA. This idea explores the fact that the schools that obtained an ISCE close to MMA but didn´t reach it, have observable and non-observable characteristics that are similar to those schools that reached the MMA. As a consequence, it

is possible to approximate the effect of the program throughout the difference between means among the change in scores of the schools that didn´t receive the incentive but were close to, and those that did receive the incentive for reaching MMA or slightly overcome it. This identification strategy is really strong, because around the MMA the assignment of the incentive is random, due to the fact that there is an estimation error in Saber test that were not considered during the program´s design. To sum up, a natural experiment is simulated around the MMA, in which some schools randomly received the incentive, and some did not.

The second strategy is defined based in the *propensity score matching method* (Caliendo & Kopeinig, 2008), that wants to compare two groups with similar visible characteristics. The strategy consists on exploiting the fact that schools that offer three different educational cycles should increase the performance in each one of them to be able to win. Therefore, schools that improved in Primary and are winners of the program, and schools that improved in Primary but didn´t win are compared. The assumption that validates this strategy, is that the performance in Middle and Secondary cycles is exogenous to Primary´s performance. In this sense, the situation presented above, would simulate an experiment in which the incentive is given randomly.

The evaluation of the program through the first strategy, is developed for the Primary school cycle, because in this scenario, by not considering the estimation error in the program´s design, it opens up the possibility to consider schools that improved their scores and for arbitrary reasons didn´t win the incentive. Therefore, in the first strategy, only the incentive effect in Primary would be evaluated, taking as a dependent variable the change in scores in Saber test applied in third and fifth grade, and only in math and language areas. Also, it is important to state that the estimations are made separately for each grade and area, with the purpose of researching if the impact varies with respect to these variables. In contrast, the second strategy can only be implemented for schools with three educational cycles, because is in this group that is possible to identify the improvement of school establishments in Primary that didn´t receive the incentive for not improving also in the other two school cycles. In this sense, by maintaining the assumption of exogeneity between Primary and other school cycles, the non-beneficiaries who improved in Primary education can be compared with the beneficiaries.

Table number 1, shows the number of schools that participated in both versions of the program, according to the offered school cycles: around 25% of the participant schools only offered primary cycle; in both versions of the program, the beneficiaries represented the 18% of the participants; in the second version of the program more than twice as many schools participated, because the eligibility criterion was not restricted to participation in PTA or JU (República de Colombia, 2017). Finally, the probability of winning the incentive, is higher in primary because the schools that just offer this cycle only have to overcome a single MMA to become beneficiaries. This last conclusion is expected, considering that in the two versions of the program the percentage of beneficiaries that only offered primary were around the 50%, while in the rest of the schools this percentage did not exceed 23%.

TABLE 1. Number of participant schools and beneficiaries

	2015		2016	
	Beneficiaries	**Non beneficiaries**	**Beneficiaries**	**Non beneficiaries**
Primary	207	160	934	1063
Primary and Middle school	55	275	135	447
Three educational cycles	451	2772	438	5188
Total	**713**	**3207**	**1507**	**6698**

Source: Icfes y MEN. Calculated by authors.

First identification strategy

Table number 2 shows the differences between means among the beneficiaries and non beneficiaries for the response variable, which is the change in the average score obtained by the schools in the Saber test, applied in third and fifth grade, and in the two mentioned areas. There are three groups of differences: for all participants, for those who were a point away of winning (MMA), and for those that were around half a point to

win. Besides, in the first version of the program, differences between means in the immediate year of implementation after the launch of the program (2015-2016), and in the second year (2016-2017) are shown. For the participants in the second version, same periods are shown, but the first one of them does not show an evaluation, because the construction of the differences between means, in this case, must be positive.

In the change observed between 2015 and 2016, the analysis with all contestants reveals that the scores of non-beneficiaries increased more than the ones reached by beneficiaries, and that these differences are significant in all grades and tested areas. When considering only the schools that were on the limit to win, which are precisely the ones who allow us to estimate the effects of the program, we find a reduction in the magnitude of the difference. But all differences, except the ones in fifth grade, when using the half point distance interval, remain negative and significant at 95% of confidence. In the second year after the launch of the program, differences between means resulted non-significate in all grades and areas tested, both in the sub-sample schools located close to the limit to win, and in all participants. In general terms, these results suggest that in the first year of implementation, the incentive diminished primary student´s performance, and during the second year it had no effect at all.

The bottom panel, Table 2, shows the results for participants in 2016. In this case, the analysis for schools close to the limit exhibits that one period after the program (2016-2017), non beneficiary schools improved more than beneficiaries. With the distance-interval of one point, differences are significant at 95% of confidence for third grade in both areas, and for fifth grade in language. In contrast, the analysis done with half a point of distance, only displays significant differences at 90% confidence for third grade, but just in language.

Finally, it is important to highlight that due to the chosen identification strategy, the estimated effect is local, meaning that it is only valid for schools that are close to the limit. This could only be applied to the total of beneficiaries, under the assumption that there is a similarity between the schools that received the incentive with ease, and those that received it with a small margin of distance from the limit.

TABLE 2. Differences between means among beneficiaries and non-beneficiaries

			Change in scores (2016-2015)						Change in scores (2017-2016)					
	Grade	Area	Beneficiaries		Non beneficiaries		Difference	Value P	Beneficiaries		Non beneficiaries		Difference	Value P
			Obs.	Media	Obs.	Media			Obs.	Media	Obs.	Media		
2015														
All participants	3°	Language	140	10.11	102	40.25	-30.14	0.0000	140	-0.85	107	0.61	-1.46	0.7618
		Mathematics	139	9.59	106	41.32	-31.73	0.0000	146	-13.53	110	-4.81	-8.73	0.1014
	5°	Language	151	13.19	112	34.99	-21.80	0.0000	155	-3.37	117	-1.94	-1.43	0.7225
		Mathematics	144	8.81	115	32.02	-23.21	0.0001	149	-10.42	120	-8.76	-1.66	0.7131
Beneficiaries and non-beneficiaries closed to the limit (1 point to ISCE)	3°	Language	67	22.06	68	46.41	-24.35	0.0021	69	2.49	73	-0.45	2.94	0.6377
		Mathematics	69	27.20	72	45.65	-18.45	0.0353	72	-13.74	74	-4.35	-9.38	0.1750
	5°	Language	75	23.71	75	37.17	-13.47	0.0296	80	-0.50	78	0.12	-0.62	0.9087
		Mathematics	75	14.27	77	33.61	-19.34	0.0051	79	-7.86	80	-7.09	-0.77	0.8893
Beneficiaries and non-beneficiaries closed to the limit (0.5 points to ISCE)	3°	Language	32	19.00	42	47.90	-28.90	0.0124	31	-0.87	43	1.88	-2.75	0.7324
		Mathematics	32	21.22	46	50.52	-29.30	0.0097	32	-12.19	44	-2.59	-9.60	0.2612
	5°	Language	35	27.11	45	39.00	-11.89	0.1938	35	0.00	47	-2.09	2.09	0.7920
		Mathematics	36	17.25	45	35.04	-17.79	0.0859	35	-9.69	47	-6.38	-3.30	0.6539
2016														
All participants	3°	Language	258	39.00	360	-8.53	47.54	0.0000	282	-9.62	377	11.01	-20.62	0.0000
		Mathematics	274	40.10	364	-16.23	56.33	0.0000	299	-18.70	374	5.03	-23.73	0.0000
	5°	Language	303	35.57	416	-0.54	36.11	0.0000	317	-5.91	422	9.01	-14.92	0.0000
		Mathematics	301	30.84	405	-11.04	41.88	0.0000	320	-13.69	418	1.42	-15.10	0.0000
Beneficiaries and non-beneficiaries closed to the limit (1 point to ISCE)	3°	Language	85	28.00	220	-4.56	32.56	0.0000	88	2.41	227	10.89	-8.49	0.0467
		Mathematics	88	27.18	228	-11.54	38.72	0.0000	96	-5.56	228	4.28	-9.84	0.0346
	5°	Language	109	27.79	255	1.94	25.85	0.0000	105	1.60	253	10.63	-9.03	0.0027
		Mathematics	109	19.52	247	-7.40	26.93	0.0000	106	-3.15	250	2.28	-5.43	0.1540
Beneficiaries and non-beneficiaries closed to the limit (0.5 points to ISCE)	3°	Language	44	22.34	119	-1.51	23.85	0.0003	44	0.98	125	11.48	-10.50	0.0979
		Mathematics	45	21.96	128	-11.78	33.74	0.0000	48	-8.88	127	4.19	-13.06	0.0648
	5°	Language	53	25.43	142	0.61	24.83	0.0000	52	3.37	141	7.96	-4.59	0.2899
		Mathematics	51	17.41	136	-8.41	25.82	0.0000	51	-3.67	140	3.70	-7.37	0.1540

Source: Icfes y MEN. Calculated by authors.

As an exercise, Table number 3 shows linear regressions for schools close to the limit (MMA). The results of one regression that includes only as an independent variable a dummy that indicates if the school is a beneficiary, would be equivalent to the exercise of the difference between means presented previously. With this in consideration, and with the purpose of validating if the estimation of the effects through the difference between means changes when adding other school characteristics, in the following regressions we include as controls: the area where the school is located (urban or rural), number of tested students and the certified territorial entity (ETC)[10] to which the school is ascribed (specifications (2) y (4)). The last characteristic is considered when including categorical variables for the 95 ETC of the country. Besides, in these specifications, an index built with the average of all obtained scores in grades and tested areas is used. As it is shown in Table number 3, the addition of this controls do not generate changes in the results found through the differences between means. In addition, for beneficiaries and non- beneficiaries, and in both versions of the program, the controls show that the area and/or the number of tested students has no influence whatsoever over the index change.

(10) In spanish: Entidad Territorial Certificada.

TABLE 3. Linear Regressions

	Difference in Scores (2016-2015)				Difference in scores (2017-2016)			
	Proximity to limit				Proximity to limit			
	One point of ISCE		Half point of ISCE		One point of ISCE		Half point of ISCE	
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Beneficiaries 2015 (=1)	**-14.85*** **	**-18.01*** **	**-21.15*** **	**-16.21***	**-3.393**	**-3.288**	**-7.026**	**-9.912**
	(-2.61)	(-3.17)	(-2.67)	(-1.96)	(-0.73)	(-0.60)	(-1.19)	(-1.32)
Urban area (=1)	**13.37**	**21.33**	**12.68**	**20.69**	**-3.417**	**-4.805**	**-4.964**	**-10.01**
	(0.72)	(1.20)	(0.67)	(1.14)	(-0.23)	(-0.28)	(-0.36)	(-0.62)
Number of tested	**0.0588**	**0.0624**	**0.0771**	**-0.0857**	**0.00183**	**-0.223**	**0.215**	**0.213**
	(0.36)	(0.30)	(0.32)	(-0.33)	(0.01)	(-1.10)	(1.23)	(0.87)
Constant	**35.36*** **	**-87.43*** **	**38.60*** **	**70.75** **	**-2.196**	**-11.51**	**-5.869**	**-15.44**
	(6.65)	(-2.73)	(5.33)	(2.25)	(-0.50)	(-0.38)	(-1.10)	(-0.55)
Fixed effects of region		Yes		Yes		Yes		Yes
N	**165**	**165**	**89**	**89**	**164**	**164**	**85**	**85**
R2	0.0443	0.400	0.0812	0.550	0.00363	0.157	0.0306	0.304
R2 Adjusted	0.0265	0.249	0.0488	0.341	-0.0151	-0.0572	-0.00528	-0.0444
F	2.489	2.645	2.504	2.624	0.194	0.733	0.853	0.872
	(1)	**(2)**	**(3)**	**(4)**	**(1)**	**(2)**	**(3)**	**(4)**
Beneficiaries 2016 (=1)	**28.32*** **	**25.96*** **	**27.09*** **	**24.52*** **	**-9.067*** **	**-9.962*** **	**-11.90** **	**-13.10** **
	(8.49)	(7.44)	(5.95)	(5.05)	(-3.03)	(-3.11)	(-2.55)	(-2.51)
Urban area (=1)	**4.727**	**9.916**	**-7.328**	**-1.865**	**-10.48**	**-6.353**	**-7.813**	**-12.01**
	(0.56)	(0.87)	(-0.71)	(-0.11)	(-1.41)	(-0.62)	(-0.75)	(-0.66)
Number of tested	**0.0625**	**-0.0154**	**0.0957**	**0.0180**	**0.0564**	**0.0110**	**0.0930**	**0.0710**
	(0.85)	(-0.13)	(0.80)	(0.11)	(0.86)	(0.10)	(0.77)	(0.40)
Constant	**-6.683*** **	**-46.14**	**-6.895** **	**0.620**	**5.655*** **	**-9.782**	**4.870**	**31.52**
	(-2.76)	(-1.44)	(-1.98)	(0.02)	(2.64)	(-0.36)	(1.39)	(0.94)
Fixed effects of region		Yes		Yes		Yes		Yes
N	**401**	**401**	**208**	**208**	**393**	**393**	**207**	**207**
R2	0.157	0.265	0.151	0.345	0.0306	0.130	0.0356	0.174
R2 Adjusted	0.150	0.182	0.139	0.208	0.0231	0.0253	0.0214	-0.00724
F	24.61	3.164	12.09	2.506	4.086	1.242	2.498	0.960

Statistics t in parenthesis.

* p<0.1, ** p<0.05, *** p<0.01

Received: 18-09-2018 Accepted: 26-04-2019

Second Identification Strategy

The second strategy identifies the effect of the incentive with a different methodology and in different schools. In this case, the impact of the incentive in three educational cycle schools is evaluated. The beneficiaries are defined as the schools that improved in Primary, Middle and High School. On the other hand, non-beneficiaries are defined, as the schools that improved in Primary but not in Middle and High School. This strategy exploits the fact that what happens in Middle and High School affects the condition of the incentive´s winner, even though the educational levels are exogenous to what happens in Primary cycle.

Table number 4 shows the differences between means among the selected schools for the second identification strategy. Although estimates are made for a group different from the one used in the previous strategy, the findings are similar. Amongst the participants in the first version, a year after the program we can observe that in average, non-beneficiaries improved more than beneficiaries, and that two years later we don't find any significant difference. For second year participants, results showed that a year after the program, non-beneficiaries in average, improved more than beneficiaries.

Received: 18-09-2018 Accepted: 26-04-2019

TABLE 4. Differences between beneficiaries and non-beneficiaries.

Grade	Area	Score Change (2016-2015)						Score Change (2017-2016)					
		Beneficiaries		Non Beneficiaries		Difference	Value P	Beneficiaries		Non Beneficiaries		Difference	Value P
		Obs.	Mean	Obs.	Mean			Obs.	Mean	Obs.	Mean		
2015													
3°	Language	436	11.74	1399	25.31	-13.56	0.0000	429	-6.05	1395	-3.49	-2.56	0.0965
	Mathematics	434	13.83	1412	30.34	-16.51	0.0000	430	-11.37	1405	-8.09	-3.28	0.0687
5°	Language	442	16.36	1436	28.59	-12.23	0.0000	434	-0.98	1421	-3.36	2.38	0.0436
	Mathematics	442	7.43	1430	19.85	-12.42	0.0000	434	-8.99	1419	-10.12	1.13	0.4166
2016													
3°	Language	411	34.58	2286	-3.21	37.79	0.0000	412	-10.14	2300	2.54	-12.68	0.0000
	Mathematics	415	38.19	2286	-4.96	43.15	0.0000	416	-17.55	2299	1.30	-18.85	0.0000
5°	Language	429	33.54	2369	8.04	25.50	0.0000	426	-6.79	2360	2.23	-9.03	0.0000
	Mathematics	427	25.02	2361	-5.02	30.03	0.0000	424	-15.29	2353	-2.14	-13.15	0.0000

Source: Icfes y MEN. Own calculations.

With the purpose of observing if there were changes in the presented findings when including control zones, number of tested and fixed effects of ETC, Table number 5 presents the results of a linear regression with different specifications. For each version of the program, three specifications are presented by group of comparison: Specification (1) controls by zone and number of people tested. Specification (2) includes fixed effects of the ETC, and specification (3) narrows the sample to schools that have a low probability of becoming beneficiaries due to the three former variables[11]. In this context, the last specification tries to minimize the differences between beneficiaries and non-beneficiaries that improved in primary by selecting schools that share the three characteristics included as controls.

(11) The probability is estimated with a Logit regression that has as a dependent variable a categorical variable that assumes the value of one if the school is beneficiary and zero if it's not. The dependent variables are: zone, number of people tested and fixed effects. Specification number 3 only includes schools with a prediction lower than 0.2.

TABLE 5. Linear Regressions

	Difference in scores (2016-2015)			Difference in scores (2017-2016)		
	(1)	(2)	(3)	(1)	(2)	(3)
Beneficiaries 2015 (=1)	**-16.73*** **	**-16.57*** **	**-16.11*** **	**2.078**	**2.591**	**4.436**
	(-14.11)	(-14.48)	(-9.37)	(0.87)	(0.33)	(0.63)
Urban area (=1)	**-2.090***	**-0.769**		**-2.109** **	**-1.092**	
	(-1.87)	(-0.67)		(-2.02)	(-0.98)	
Number of tested	**-0.0483*** **	**-0.0434*** **		**-0.0207***	**-0.0317** **	
	(-3.89)	(-3.39)		(-1.79)	(-2.56)	
Constant	**33.78*** **	**26.55*** **	**30.32*** **	**-6.507*** **	**1.437**	**-8.338*** **
	(39.97)	(3.51)	(46.85)	(-8.21)	(0.20)	(-13.24)
Fixed effects of region		Yes			Yes	
N	**2372**	**2372**	**1383**	**2338**	**2338**	**1360**
R2	0.0935	0.221	0.0598	0.00815	0.0817	0.00507
R2 Adjusted	0.0923	0.191	0.0591	0.00687	0.0449	0.00434
F	81.39	7.203	87.77	6.39	2.22	6.927
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
Beneficiaries 2016 (=1)	**33.16*** **	**30.58*** **	**32.07*** **	**-19.48*** **	**-20.28*** **	**-19.74*** **
	(27.44)	(25.54)	(19.6)	(-22.80)	(-23.39)	(-16.91)
Urban area (=1)	**-2.319** **	**-0.953**		**-4.243*** **	**-3.042*** **	
	(-2.31)	(-0.91)		(-5.97)	(-4.02)	
Number of tested	**0.00225**	**-0.0142**		**-0.0213*** **	**-0.0210** **	
	(0.19)	(-1.12)		(-2.60)	(-2.30)	
Constant	**0.757**	**10.45**	**-1.372** **	**10.89*** **	**13.99** **	**7.570*** **
	(1.04)	(1.24)	(-2.52)	(21.31)	(2.31)	(19.58)
Fixed effects of region						
N	**3035**	**3035**	**2142**	**3005**	**3005**	**2125**
R2	0.2	0.286	0.152	0.173	0.228	0.119
R2 Adjusted	0.199	0.263	0.152	0.172	0.202	0.118
F	253	12.15	384.3	209.7	8.85	285.9

Statistic t in parenthesis.
* p<0.1, ** p<0.05, *** p<0.01
Source: Icfes y MEN. Own Calculation

The findings of the linear regressions that consider the characteristics of the schools, show very similar results to the ones found in the differences between means. In the group of participants of 2015, beneficiary's improvement was lower than non- beneficiaries in the year that followed the program, and during the second year there are no substantial differences. Furthermore, a year after the program was launched in 2016, we can observe that, in average, non-beneficiaries showed better results. At last, overall it can be seen that urban schools improved less than the countryside ones, and schools with a smaller number of students displayed better results, which is consistent with the results found in the literature about the subject.

Finally, in the same way that happens with the first strategy, the incentive effect estimated through the second strategy is local. Although, because both strategies were implemented in different school groups, it is possible to infer that the findings shown above could be applied to a good portion of the population.

Discussion and Conclusion

Identifying the added value to teacher´s efforts is not an easy task due to the confluence of multiple factors that affect the learning process such as: peer effect (Hanushek et al., 2003; Lin, 2010; Zimmerman, 2003), socioeconomic conditions and parent involvement (Fan & Chen, 2001; White, 1982). Hence, is necessary to use econometric methods to try to decipher which part of the performance is a direct consequence of the teacher´s effort, and not to other observable and non-observable characteristics related to the learning process. In this sense, it would be incorrect to evaluate the economic incentives program´s impact through the variations in the Saber test scores, because is not reasonable to assert that the changes in performance are exclusively related to the incentive. To minimize the problems of this approach, in this article we used the design of the program to build sub samples of beneficiaries and non-beneficiaries with similar characteristics, so we would be able to state that, overall, the only difference between the subsamples is the incentive, and, therefore the difference between means would approximate the effect of the program.

Having the goal of minimizing the problems already described, two main strategies were used in this article to estimate the local effect of

the program, or in other words, the effect over a specific subsample of participants. Altogether, through the *discontinuity regression method* in the first strategy, beneficiary and non-beneficiary schools that were close to the limit for winning were used. They were chosen because among these establishments is possible to argue that the programs assignment is random due to the existence of an estimation error in Saber test scores, that was not considered in the program´s design. The second strategy is based on the propensity *score matching method* and was implemented under the assumption that there is a similarity between schools that improved their ISCE in primary (non-beneficiaries) and those who did well in all three cycles (beneficiaries). Although the results gathered through these strategies are only valid in the subsamples used, and under the assumptions already mentioned; the fact that the subsamples applied in each strategy are different, validates the results for a greater percentage of participants. Overall, the results discovered are counterintuitive, because they indicate that the incentive had no positive effect over the student's performance. In fact, during the first year of implementation non-beneficiaries, in average, showed improvement over beneficiaries, and during the second year of implementation there are no significant differences. Up next, we will discuss possible mechanisms that could help comprehend our findings.

One of the possible explanations is related to *free rider problem***.** Because the incentive is assigned based on the scores of Saber, the teachers that are not involved in the grades and tested areas, receive no incentives to potentiate their efforts, because they can obtain them effortless. Concerning this aspect, Biener et al. (2018) point out that an incentive delivered through a group payment scheme works if the agents care about the results of their peers. Consequently, as Imberman & Lovenheim (2015) propose, the incentive effect is amplified, if the proportion of students that determine the assignment of the incentive and that are the responsibility of the teacher is bigger, in this sense the contribution of each teacher to obtain the incentive is greater. The alternative for group assignments is individual assignment, which, in spite of solving *free rider problem,* has the disadvantage of omitting pedagogical approaches that defend the existence of complementarities between grades and areas in the educational process.

Another aspect to consider is the program´s temporality. It may be argued that the absence in important differences between beneficiaries and

Received: 18-09-2018 Accepted: 26-04-2019

non beneficiaries, is related to the fact that all eligible schools had the opportunity to earn the incentives, so, the changes in teachers behaviors may be more related to the existence of the program than to the incentive itself (Shifrer et al., 2017). In this scenario, the evaluation presented in this article may not be adequate because it didn´t considered the schools that had the possibility to earn the incentive. In addition, it must be considered as a main factor, the little amount of time elapsed between the launching of the program and the test delivery that lead to the calculation of the ISCE published in 2016. The program was announced at the beginning of July 2015, and only three months later was the test that determined the incentive applied. This situation does not consider that in such a small period is unlikely that the changes in teacher´s behavior generate changes in students´ performance. Taking this into account, the revision of the theoretical approaches does not reveal any reasons that can explain the differential behavior in favor of non-beneficiaries. This result may be explained with the argument that, beneficiaries, more than displaying an effort to their teachers, are showing a positive random shock, that is compensated the next year with a decrease. In this case, beneficiaries and non-beneficiaries, would be the result of the separation between schools that received a positive and negative shock in this particular year. Two years later there will be no difference.

Is important to highlight that the program may present different effects on other measures not studied in this article, such as teacher absenteeism and school dropout. To sum up, we conclude accentuating the importance of keeping in mind studies like this one, to make public policy decisions associated to the design and implementation of programs that tend to improve the quality of education through economic incentives for teachers. For later studies, we suggest to investigate the pertinence of incentives delivered individually, and the use of measures that demonstrate changes in time.

Bibliographical References

Allan, B. M., & Fryer, R. G. (2011). The power and pitfalls of education incentives. *The Hamilton Project*, (September), 1–36. Retrieved from http://www.brookings-tsinghua.cn/~/media/Files/rc/papers/2011/09_incentives_fryer_allen/092011_incentives_fryer_allen_appendix.pdf

Baker, G. (2002). Distortion and Risk in Optimal Incentive Contracts. *Journal of Human*
Resources, 37(4): 728–51.

Biener, C., Eling, M., Landmann, A., & Pradhan, S. (2018). Can group incentives alleviate moral hazard? The role of pro-social preferences. *European Economic Review*, *101*, 230–249. https://doi.org/10.1016/j.euroecorev.2017.10.004

Buddin, R., & Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, *66*(2), 103–115. https://doi.org/10.1016/j.jue.2009.05.001

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31–72. https://doi.org/10.1111/j.1467-6419.2007.00527.x

Conrads, J., Irlenbusch, B., Rilke, R. M., & Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology*, *34*, 1–7. https://doi.org/10.1016/j.joep.2012.10.011

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, *18*(1), 105–115.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). The undermining effect is a reality after all—Extrinsic rewards, task interest, and self-determination: Reply to Eisenberger, Pierce, and Cameron (1999) and Lepper, Henderlong, and Gingras (1999). *Psychological Bulletin*, *125*(6), 692–700. https://doi.org/10.1037/0033-2909.125.6.692

Duflo, E., Rema, H. and Stephen P.R. (2012). Incentives Work: Getting Teachers to Come to School. *American Economic Review* 102(4): 1241-1278.

Eijkenaar, F. (2013). Key issues in the design of pay for performance programs. *The European Journal of Health Economics*, *14*(1), 117–131. https://doi.org/10.1007/sl0198-011-0347-6

Fan, X., & Chen, M. (2001). Parental Involvement and Studnet's Academic Achievement: A Meta-Analysis. *Educational Psychology Review*, *13*(1), 1–22. https://doi.org/10.1023/A:1009048817385

Fryer, R. (2013). Teacher Incentives and Student Achievement: Evidence from New York City Public Schools, *31*(2), 373–407. https://doi.org/10.3386/w16850

Fryer, R., Levitt, S., List, J., & Sadoff, S. (2012). Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment. https://doi.org/10.3386/w18237

Glewwe, P., Ilias, N., Kremer, M., American, S., Journal, E., Economics, A., & July, N. (2010). Teacher Incentives. *American Economic Journal: Applied Economics*, *2*(3), 205–227. https://doi.org/http://dx.doi.org/10.1257/app.2.3.205

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives*, *25*(4), 191–210. https://doi.org/10.1257/jep.25.4.191

Goodman, S. F., & Turner, L. J. (2013). The Design of Teacher Incentive Pay and Educational Outcomes: Evidence from the New York City Bonus Program. *Journal of Labor Economics*, *31*(2), 409–420. https://doi.org/10.1086/668676

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, *18*(5), 527–544. https://doi.org/10.1002/jae.741

Hanushek, E. A., & Woessmann, L. (2011). Overview of the symposium on performance pay for teachers. *Economics of Education Review*, *30*(3), 391–393. https://doi.org/10.1016/j.econedurev.2010.12.005

Holmstrom, B., & Milgrom, P. (1991). Multitask Principal-Agent Analyses : Incentive Contracts , Asset Ownership , and Job Design. *Journal of Law, Economincs and Organization*, *7*(Special Issue), 24–52. https://doi.org/10.1093/jleo/7.special_issue.24

Hudson, S. (2010). The effects of performance-based teacher pay on student achievement. *SIEPR Discussion Papers*, *94305*(09). Retrieved from http://cell.uindy.edu/docs/TAP/EffectsPerfBasedTeacherPayStudentAchievement.pdf

Icfes. (2016). ISCE: Guía Metodológica. *Boletín Saber En Breve*.

Icfes. (2017). La tercera versión del índice sintético de la calidad educativa (ISCE 2017). *Boletín Saber En Breve*.

Imberman, S. A., & Lovenheim, M. F. (2015). Incentive strength and teacher productivity: evidence from a group-based teacher incentive pay system. *Review of Economics and Statistics*, *97*(2), 364–386. https://doi.org/10.1162/REST

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, *3*(4), 305–360. https://doi.org/10.1016/0304-405X(76)90026-X

Lavy, V. (2002). Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement. *Journal of Political Economy*, *110*(6), 1286–1317. https://doi.org/10.1086/342810

Lavy, V. (2009). Performance pay and teachers' effort, productivity, and grading ethics. *American Economic Review*, *99*(5), 1979–2011. https://doi.org/10.1257/aer.99.5.1979

Lee, D. S., & Lemieux, T. (2010). in Economics. *Journal of Economic Literature*, *20*(1), 281–355. https://doi.org/10.1257/jel.48.2.281

Lin, X. (2010). Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables. *Journal of Labor Economics*, *28*(4), 825–860. https://doi.org/10.1086/653506

Marsh, J. A., Springer, M. G., McCaffrey, D. F., Yuan, K., Scott, E., Koppich, J., … Peng, A. (Xiao). (2011). *A Big Apple for Educators New York City's Experiment with Schoolwide Performance Bonuses*. RAND Corporation.

Martins, P. S. (2009). Individual Teacher Incentives, Student Achievement, and Grade Inflation. *Applied Economics*, (4051), 43. Retrieved from http://ftp.iza.org/dp4051.pdf

MEN (2015). Comunicado de prensa. Obtenido de: https://www.mineducacion.gov.co/cvn/1665/w3-article-351705.html

Muralidharan, K., & Sundararaman, V. (2011). Teacher opinions on performance pay: Evidence from India. *Economics of Education Review*, *30*(3), 394–403. https://doi.org/10.1016/j.econedurev.2011.02.001

OCDE. (2016). *La Educación en Colombia: Revisión de políticas nacionales de educación*. Paris: OCDE. https://doi.org/10.1787/9789264250604-en

República de Colombia. (2017). Decreto N.o 1577, Por el cual se modifican los artículos 2.3.1.6.6.2 y 2.3.1.6.6.3 del Decreto 1075 de 2015. Bogotá: Diario Oficial.

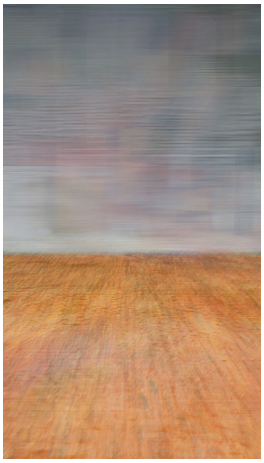
Ryan, R. M. (1982). Control and Information in the Intrapersonal Sphere: An Extension of Cognitive Evaluation Theory. *Journal of Personality and Social Psychology*, *43*(3), 450–461.

Shifrer, D., Turley, R. L., & Heard, H. (2017). Do Teacher Financial Awards Improve Teacher Retention and Student Achievement in an Urban Disadvantaged School District? *American Educational Research Journal*, *54*(6), 1117–1153. https://doi.org/10.3102/0002831217716540

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, *91*(3), 461–481. https://doi.org/10.1037/0033-2909.91.3.461

Zimmerman, D. J. (2003). Peer Effects in Academic Outcomes: Evidence From a Natural Experiment. *The Review of Economics and Statistics*, *85*(1), 9–23. https://doi.org/10.1162/003465303762687677

Contact address: Silvana Godoy Mateus. Instituto Colombiano de Bienestar Familiar, ICBF. Bogotá. Calle 23, nº 13-24, apartamento 503, Botogá. E-mail: silvanagodoymateus@gmail.com



Reviews

Reviews

Ruiz-Corbella, M. y García-Gutiérrez, J. (Eds.) (2019). Aprendizaje-Servicio. Los retos de la Evaluación. Madrid: Narcea. 206 pp. ISBN 978-84-277-2531-7

The university, regarded as an institution, is both a current witness and an active accomplice of the changes taking place at an increasingly vertiginous pace which, in turn, demand an equally urgent response not only capable of interpreting these changes, but which can even anticipate them. Thus, as a result of this interaction with the feedback force, it remains in constant motion, as a living being, considering at least two parameters: on the one hand, those parameters where the political-administrative engineering has been devised around some key concepts such as the so-called University Social Responsibility (USR), sustainability and educational innovation, as well as inclusion, gender perspective and many others; and, on the other hand, those essential parameters that continuously question the social, economic, political, ideological and axiological contexts from which it draws its meaning, now encapsulated in the remarkable concept of transference.

This scenario contemplates university teaching and learning models that are in a way able to meet the expectations issuing from those models of the university that transcend the classical and nowadays surpassed diatribe between the academic and professionalizing roles of the university. Furthering our metaphor, this is the moment in which Service Learning comes into play as the main protagonist. Known by its acronym, ApS [SL], as well as other names, such as Learning and Service, mainly used in the Latin-American milieu as referenced in this book, this pedagogical methodology, whose precursors date back to the end of the nineteenth century, "has come to stay", just as I have already stated before. In this way I reconfirm that, after its emergence, we have witnessed its remarkable expansion in the last third of the past century, in terms of research, actual teaching and teaching innovation, as well

as its continuous institutionalization in the universities throughout the globe.

Delving more deeply into this idea, the current state of maturity has lately crystalized in research tendencies that are intimately linked to teaching. Here I would like to refer to three such tendencies, stating in this way the purpose of this book. First, there is an outpouring of meta-analyses that systematize and categorize the extant investigations according to different criteria, aiding the charting of the dominant trends in worldwide investigations pertaining to the matter at hand. Second, the review of epistemologies enables the exegesis of theoretical models to draw from, as in the case of the current allusion to Critical Service Learning as a pedagogical-philosophical construct that points in a more transformative way to social justice not merely as a principle but also as evidence, thus enabling more authentic relationships that shatter the unequal roles of power in a practical self-knowledge regenerated in its fidelity.

Finally, the third trend, inexcusable and logical given the demand it has generated and still generates, is the one found in this book: time has come for evaluation in Learning Service. Their editors lavish us with a careful selection of compulsory references for all who investigate, reflect upon and *think and feel* Learning Service: it comprises the theoretical foundations required for the attainment of *for what* answers (for example, the model of competencies and quality indicator), as well as *why* answers (the competency and civic commitment) and *how* answers, for which some chapters aim at teaching the selection of indicators for impact measuring, recommendations for the measurement of the innovative component and social transference, as well as samples of precise research instances (lessons of *know how*), and an authentic tool kit, in the most authentic sense, as a starting guide into the tools of measurement and evaluation of the ApS: participative and self-managed evaluations, field journals and rubrics, as well as their institutionalization.

In sum, this manual contains all questions regarding evaluation in the form of a careful edition informed by the expertise of the authors that endorse this book.

María-Jesús Martínez-Usarralde

Torrego, J. C., & Monge, C. (Coord.). (2018). Inclusión educativa y aprendizaje cooperativo. Madrid: Síntesis. 300 pp. ISBN: 9788491712503

The construction of a fairer society cannot be achieved without an inclusive educational system. At the same time, schools can hardly move towards inclusion if society does not go hand in hand with them. Although inclusive postulates seem to have made an impression on the discourse of a large part of the educational community, multiple barriers prevent many of the advances supported by research from being widely reflected in practice. In the book *Inclusión educativa y aprendizaje cooperativo,* the relationship between these two concepts is addressed. Why is cooperative learning necessary to move towards a more and more inclusive education?

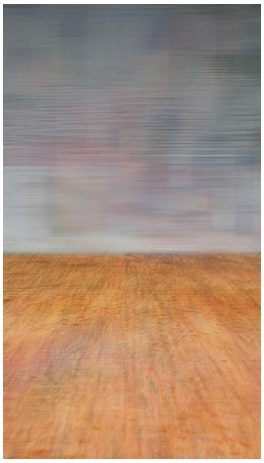
Cooperative learning and inclusion share conceptions and values, starting from the recognition and celebration of differences. Recognising the differences between people is not only fair for those groups at greater risk of exclusion, but for the whole society, because diversity is inherent to human beings. Celebrating differences between people should be the starting point to change the way we understand the teaching and learning practices: diversity and diversification, as opposed to uniformity and standardisation. Inclusive education struggles for the presence, participation and achievement of all students, and focuses on how to change contextual conditions so that individual differences do not become obstacles to learning. Cooperative learning finds in the differences between students the source of opportunities for everyone to learn by interacting with each other, and focuses on structuring that interaction to ensure the presence, participation and achievement of all of them.

Moving from conceptions to practices is a complex process, because the path to inclusion requires an explicit, strategic and systemic effort (Chapter 1). The biggest hurdle is found in some equivocal or incomplete conceptions that coexist in our discourse. One of the most widespread misconceptions, suggested by the practices of many schools, is that having students work together is enough to foster inclusion and to learn from each other. One of the premises of cooperative learning is the need to explicitly organise interaction among students, to generate relationships of social interdependence (Chapter 5). In order for students

to learn by working together –and to learn to work together– teachers need to act as mediators to create the teams, structure the interaction, help them develop social skills, and offer them tools with the aim of progressively giving them control over the regulation of their own learning (Chapter 9). Not only will organising the interaction between students offer learning opportunities to those who receive help, but it will also generate the possibility that students learn by teaching their classmates (Chapter 6). The potential of learning by teaching, together with other measures, opens valuable options for the use of cooperative learning with high-ability students (Chapter 7). Overcoming the barriers that arise along the way requires accompanying teachers in the process of change generated by inclusion and the introduction of cooperative learning in the classroom. In the search for institutional coherence (Chapter 2), this process of change leads us to rethink the leadership of school management teams (Chapter 3), teacher training (Chapter 4) and collaborative counselling (Chapter 10).

Educational inclusion and cooperative learning are mutually justified, covered up with the ideals of justice and democracy. Ideals which, under an apparent and dangerous sense of achievement, the current situation requires us to continue reclaiming. Inclusion is an endless process of analysis and reflection on equity in the educational policies and practices. The role of research is crucial to continue illuminating the path to inclusion. Research must run away from the quantitative-qualitative dichotomy and approach mixed-methods, as well as involve the entire educational community not only as informants but also as agents of research and change in their contexts (Chapter 8). The sixteen authors participating in the book, recognised in the field of inclusive education and cooperative learning, invite us to embark on a journey towards a more inclusive society and school, along a path that we cannot walk alone.

Jesús Ribosa Martínez



Annual report 2018

Annual report 2018. *Revista de Educación*

José Luis Gaviria
Editor Jefe

This report summarizes the editorial activities of the Revista de Educación in 2018. It shows the statistical data of the articles received and published, as well as the main advances in the edition of the journal.

In this issue, we can find the list of reviewers who have evaluated articles for this period as well as the list of authors who have published articles in the journal in 2018.

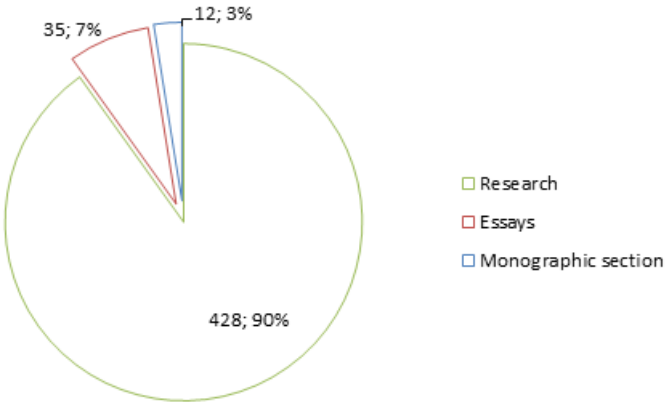
Articles received and published in 2018

Articles received by sections

The Editorial Office of our journal received a total of 475 articles throughout 2018, representing a decrease of about 8% over 2017. In figure I is shown the distribution of articles by section.

The research section still gets the largest number of originals (90%)

FIGURE I. Total number of articles received in 2018 by sections

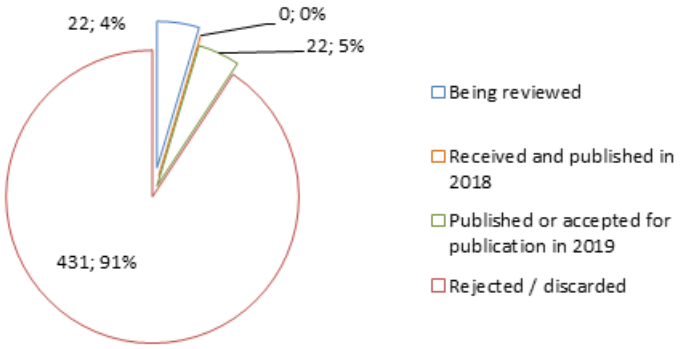


Results of the external peer-reviewed articles: Accepted and rejected articles in 2018

Of all the research articles received by the Editor, 91 % were rejected or discarded.

Of all the research papers received in 2018, 22 are still under review, 431 have been disregarded or rejected, and 22 have been published or accepted for publication in 2019. The rest of the articles published in 2018 were not received in 2018 and therefore were not counted in this computation.

FIGURE II. Progress of the articles received in 2018



In 2018 the journal published a total of 36 articles.

The following figure shows the distribution of articles published in the different sections by year (2017-2018).

FIGURE III. Comparison of published articles by section (2017-2018)

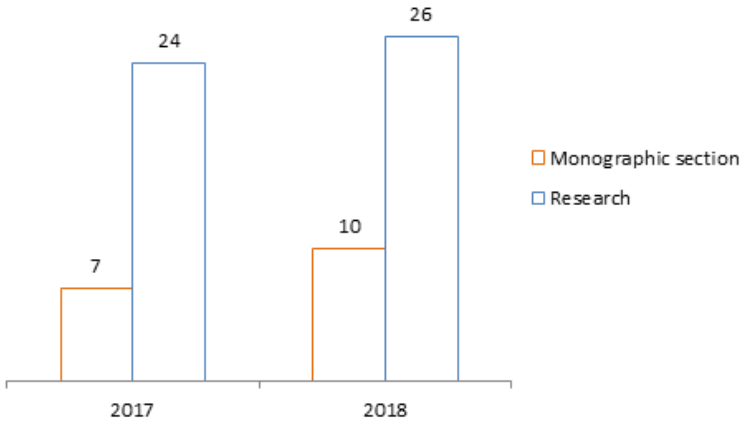
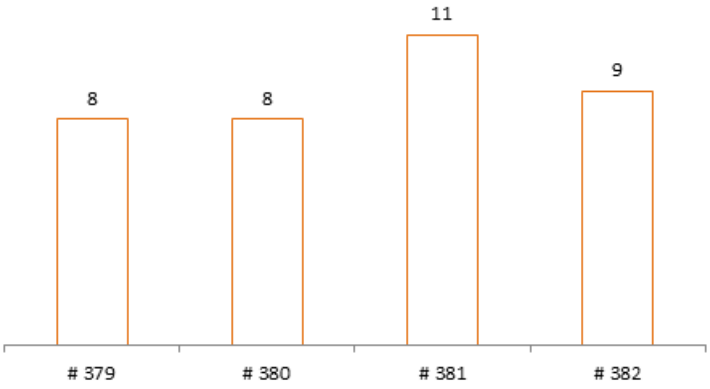


Figure IV shows the distribution of all articles published in 2018 by issue.

FIGURE IV. Articles published in each issue of the journal in 2018



The editorial policy of the journal has as a priority the publication of articles of only the highest scientific quality and the utmost general interest. This implies a reduced number of original articles published with regards to previous periods. That reduction, together with stricter selection criteria has had positive repercussions on the impact factor of the journal, as will be developed in further sections.

Topics of the monographic section in 2018

The monographic section of each issue has disappeared since number 361 onwards. Editorial policy establishes that one issue per year would include a monographic section at most, always carefully considering the special interest of the selected topic.

Publishing process: Management, revision and publication of articles

Average time between article reception and its final publication

Figure V shows the average time, in days elapsed from reception of a paper to its final publication.

Time elapsed between the reception of an original and its publication has reduced along 2018. That time depends, in general terms, on the number of articles received and, in particular, on the diligence of external reviewers.

This reduction is a primary objective of our publication, for the benefit of our journal and for the authors' benefit too.

The shorter the time between the ending of a research and its availability to the community, the better for that community and for the authors.

As for the rejected articles, it is for the best interest of the authors to know such information as soon as possible so that they can make the necessary modifications, or to find a more suitable publication mean.

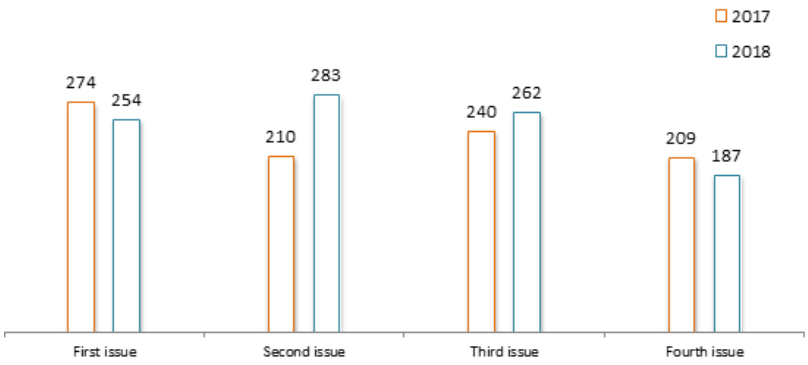
For that reason, Revista de Educación is making an effort to reduce to a minimum the processing time of articles discarded in the first review. This way the authors can look for alternatives for their manuscripts while the pressure over the reviewers of the journal is minimized, being able then to focus on articles that passed the first round.

With this objective in mind, the journal has published a list of categories of works that as a general rule, and but for cases whose exceptionality the editorial board will consider, Revista de Educación will not publish:

- Summaries of literature on a given topic
- Summaries of academic papers
- Instrument validation studies
- Evaluations of particular intervention programs
- Opinion or attitude survey results
- Reports on didactic innovations at local level
- Scientific divulgation works
- Opinion articles
- Studies in which the sample used and the method of its selection, the instruments or their technical characteristics are not clearly specified
- Works based on small or incidental samples, such as groups of students from a single school or University with little possibility of generalization

In this sense, if a work has recently been published on a given subject, it is unlikely that the same topic will be addressed again, except that the new article supposes a very relevant contribution.

FIGURE V. Average time between reception and publication of related issues (2017-2018).



2018 Revision statistics

FIGURE VI. Average time in calendar days to respond to the request for review (2018)

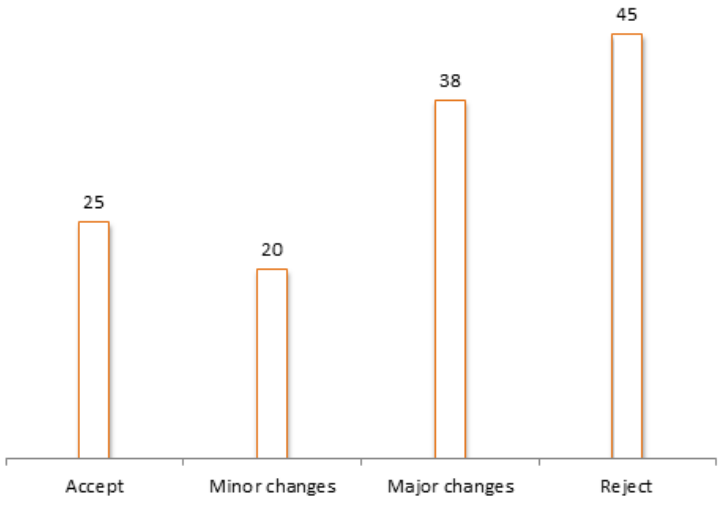
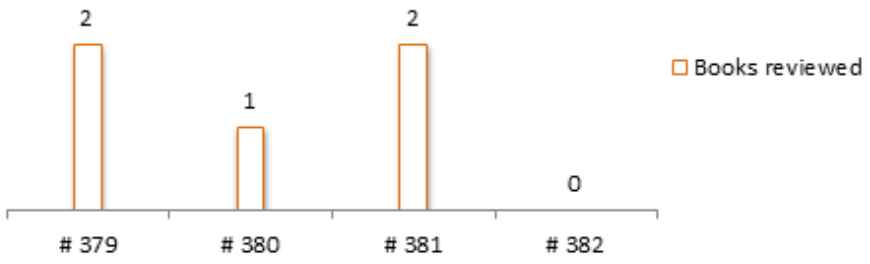


Figure VI shows that, as an average, reviewers met the term of the requested review. These are average values, implying that in some cases the process can take a significantly a longer or shorter time. As can be seen, it is precisely in the case of the rejected articles when the time involved is longer.

Dissemination of relevant works in the field of education

In 2018, 5 book reviews were published. Figure VII shows the distribution of published reviews by issue.

FIGURE VII. Books reviewed and books received in every issue published in 2018



Editorial strategy and results

Along 2018 the process initiated in 2005 has continued. This process has the objective of aligning Revista de Educación with the most demanding quality indicators for scientific journals as well as improving its impact factor, particularly in the most prestigious international databases.

Revista de Educación appears in the following sources of bibliographic documentation:

National Databases

- RESH (Revistas Españolas de Ciencias Sociales y Humanas)
- BEG (GENCAT)
- ISOC
- PSICODOC
- DIALNET
- REDINED (Red de Bases de Datos de Información Educativa)

International Databases

- Social Sciences Citation Index (SSCI)
- Social Scisearch®
- Journal Citation Reports/Social Sciences Edition
- SCOPUS (Elsevier B.V.)
- European Reference Index for the Humanities (ERIH)
- Ulrich's Periodicals Index Directory
- LATINDEX (Iberoamericana)
- Sociological Abstracts (CSA Illumina)
- PIO (Periodical Index Online, Reino Unido)
- IRESIE (México)
- ICIST (Canadá)
- HEDBIB (UNESCO-Higher Education Bibliography)
- SWETSNET (Holanda)

Platforms of journal evaluation

- SCImago Journal & Country Rank (SJR)
- CARHUS Plus+
- Matriu d'Informació per a l'Avaluació de Revistes (MIAR)
- Clasificación Integrada de Revistas Científicas (CIRC)
- Difusión y Calidad Editorial de las Revistas Españolas de Humanidades y Ciencias Sociales y Jurídicas (DICE)

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)

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 (Reino Unido)
- SUDOC, Catalogue du Système Universitaire de Documentation (Francia)
- ZDB, Zeitschriftendatenbank (Alemania)

Clarivate Analytics published, in Jun 2018, the 2017 impact factor of journals indexed in the Social Sciences Citation Index (SSCI).

Revista de Educación has an impact factor of 1.783 in JCR, occupying the position 75 of 239 in the set of journals belonging to the subject category Education & Educational Research. This indicator is based on articles published in 2016 and 2015.

In the Journal Citation Reports (JCR) published on Jun 2018, Revista de Educación has much improved its impact factor with regards to 2017. The 2017 impact factor is 1.783 while the corresponding value of 2016 was 1.185.

On 2015 the journal climbed from Q3 to Q2 on the Education & Educational Research heading. Its position has also evolved from 108th out of 235 to 75th out of 239.

Since 2010 the journal has reached its highest impact factor as well as the highest impact factor when the citation window is five years long (1.584). Similarly, the average citation life has peaked on 2017. This indicator shows us that half of the Revista de Educación articles cited on 2017 were published 8.1 years back on average.

This indicator provides an approximation to the current validity of articles published since 2010.

Despite these indicators being a positive sign of the quality of the journal, many factors affect its stability. It means we cannot only rely on those values. Comparisons with other journals on yearly basis show us a great deal of variability. That is why our goal is to keep a stable editorial line assuring a good position of the journal with regards to the consideration of the educational and scholarly community it serves.

Further information on Journal Citation Reports and the impact factor can be found at: www.accesowok.fecyt.es/jcr/

Revista de Educación is published only in electronic format, although there is an option to receive a printed version if expressly requested of the Subdirección General de Documentación y Publicaciones del Ministerio de Educación, Cultura y Deporte, which has established a printing service on demand. This same option serves subscriptions (individual and institutional), sales, and exchanges with other prestigious national and international educational journals which are included in the Education Library.

Acknowledgements

We cannot conclude this report without expressing our recognition and appreciation to all those who make Revista de Educación possible. From authors who show their appreciation by sending us their originals, to reviewers who selflessly and objectively evaluate each paper, to all the staff working for Revista de Educación who make the publication physically possible. With the help of all of them, Revista de Educación will endure in its effort to provide the best service to education and to the educational community, by allowing the fluid and vivid exchange of the results of the best scientific research in this field.

LIST OF AUTHORS, 2018

Gerardo Prieto Adánez
Gonzalo Almerich
Gonzalo Jover Olmeda
Gonzalo Sanz-Magallón Rezusta
Ignacio Berzosa Ramos
Inmaculada Asensio Muñoz
Irene Verde
Isaías Martín Ruiz
Jaleh Hassaskhah
Javier Tourón
Jesús Miguel Rodríguez-Mantilla
Jordi Díaz-Gibson
Jorge Leonardo Duarte Rodríguez
José L. Arias-Estero
José Manuel Cordero Ferrera
José María Roa
José Miguel Vílchez González
José-Miguel Nieto
Josetxu Orrantia
Juan-Manuel Escudero
Laura Matilla
Leticia Gallegos-Cázares
Leticia Williams Pinto
Lieven Verschaffel
Lourdes Meroño
Luis Lizasoain Hernández
Mª Carmen García Centeno
Mª Jesús Colmenero Ruiz
Mª José Fernández-Díaz
Mar Camacho Martí
Marco Rieckmann
María Antonia Manassero Mas
María Dolores López-Jiménez
María Dolores Martín-Lagos López
María Gil Izquierdo
María José Arroyo González
María José Cuetos Revuelta
María José García San Pedro

María José Rodríguez-Conde
María Luisa Martí Selva
María Luz Diago Egaña
María Martínez Hita
María-José González-Valenzuela
Marta Muñoz Hernández
Martha Lucia Landron
Mireia Civís Zaragoza
Miren Fernández-de-Álava
Miriam Agreda Montoro
Natalia Pinto Solano
Norberto O. Corral Blanco
Pablo Sayans-Jiménez
Patricia González González
Patricia Suárez
Philipp Christian Grollman
Piedad Mª Sahuquillo Mateo
Pilar Pineda-Herrero
Raquel Fernández-Cézar
Rosa Puertas Medina
Rosario Sánchez
Sara Álvarez Morán
Sara San Romualdo
Silvana Godoy Mateus
Susana López Solé
Susana Olmos-Migueláñez
Teresa Rivas-Moya
Tobías Martín Páez
Trinidad Suárez
Manuel M. Molina-López
Víctor Cristobal López
Víctor Valdivia Rodríguez
Ximena Dueñas Herrera

EXTERNAL REVIEWERS 2018

Aguilar-Rivera María Del Carmen
Aliaga Francisco M.
Álvarez Lires María
Angulo Rasco José Félix
Antón Pérez José Ignacio
Antúnez Marcos Serafín
Area Manuel
Arraiz Pérez Ana
Arribas Galarraga Silvia
Arroyo Resino Delia
Asensio Muñoz Inmaculada
Azpillaga Verónica
Azpillaga Verónica
Balart Pau
Ballesteros Velázquez Belén
Barbera Ortega María Del Carmen
Baró Lambias Mónica
Bartolomé Pina Antonio
Bernal Guerrero Antonio
Bernal Martínez José Mariano
Biencinto Chantal
Blanco Blanco Ángeles
Canales Antonio
Carabaña Morales Julio
Cárdenas Denham Sergio
Carpintero Elvira
Carreño Miryam
Casals Ibáñez Albert
Castilla Mesa Mª Teresa
Castro Inostroza Angela Nolfa
Castro Martínez Encarnación
Cercadillo Lis
Contreras García José Miguel
Cordón Lagares Encarnación
Corredera González Azucena
Cubo Delgado Sixto

Curiel-Marín Elvira
De Castro Hernández Carlos
De Haro Rodríguez Remedios
De La Herrán Gascón Agustín
De La Torre Cruz Manuel Jesús
De Pro Bueno Antonio
Delgado-Ponce Águeda
Díaz Pérez Flora Mª
Dietz Gunther
Domingo Segovia Jesús
Elexpuru Albizuri Itziar/Iciar
Escarabajal Arrieta María Dolores
Estepa Giménez Jesús
Expósito Casas Eva
Fernández Álvarez Miguel
Fernández Andrés María Inmaculada
Fernández Cruz Manuel
Fernández Enguita Mariano
Fernández Navas Manuel
Fernández Sierra Juan
Fernández-Mellizo María
Fidalgo Redondo Raquel
Fontana Abad Mónica
Fuentes Juán Luís
Galicia Mangas Francisco Javier
Galindo-Domínguez Héctor
Gallardo Vigil Miguel Ángel
García García Mercedes
García Gutiérrez Juan
García Laborda Jesús
García-García Mercedes
Gento Palacios Samuel
Gewerc Adriana
Gil Galván Mª Rosario
Giménez Toledo Elea
González Barbera Coral
Gonzalo Ignacio
Grad Héctor Mario

Gutiérrez-Braojos Calixto
Guzmán-Simón Fernando
Hernández Daniel
Hernández Franco Vicente
Ianos Adelina
Ibáñez Echeverría Alejandro
Iborra Cuéllar Alejandro
Ilabaca Baeza Paola
Imbernón Francisco
Jiménez Eva
Jiménez Jaen Marta Esther
Jornet Meliá Jesús Miguel
Jover Olmeda Gonzalo
Latorre Medina María José
León Carrascosa Víctor
Lliná S. Audet Xavier
López Facal Ramón V.
López González Emelina
López Martín Esther
López Melero Miguel
López Ruiz Juan Ignacio
Luque Mengibar David
Manzano Soto Nuria
Mañana Rodríguez Jorge
Marhuenda Fernando
Marí-Klose Pau
Márquez López Jorge
Martínez Ruiz M. Ángeles
Mata Benito Patricia
Mateo Andrés Joan
Miret Ines
Mohedano Fuertes María Jesús
Molto Egea Oscar
Moreno Botella Gloria
Moreno Celia
Moriyón Félix García
Muñoz Cantero Jesús Miguel
Muñoz Losa Aurora

Murillo Estepa Paulino
Murillo Torrecilla F. Javier
Navarro Asencio Enrique
Navas Julia
Ordoñez Xavier G.
Peralbo Uzquiano Manuel
Pérez Esteve Pilar
Pérez Esteve Pilar
Pérez Pérez Cruz
Portela Pruaño Antonio
Pozo Juan Ignacio
Prado-Domínguez Antonio Javier
Puertas Medina Rosa
Quiñones Jesús
Ramos Mendoza José Roberto
Rebollo Quintela Nuria
Rebollo-Catalán Ángeles
Rico Romero Luis
Rivero Gracia M. Pilar
Robledo Ramón Patricia
Romero Pérez Clara
Ruipérez Valiente José A.
Ruíz Corbella Marta
Ruiz De Miguel Covadonga
Ruiz Lázaro Judit
Salas Labayen María Rosa
Sánchez-Mesa Domingo
Sanz-Magallón Rezusta Gonzalo
Serra Josep María
Somoza Rodríguez Miguel
Souto Xosé Manuel
Suárez Ortega Magdalena
Suárez Rodríguez Jesús M.
Tiana Ferrer Alejandro
Tourón Figueroa Javier
Valenzuela Jorge
Vera Jiménez Jesús
Vila Merino Eduardo Salvador
Yáñez-Marquina Lara

Revista de Educación is a scientific publication of the Spanish Ministerio de Educación y Formación Profesional. Founded in 1940, with the title *'Revista de Educación'* since 1952, it has been an exceptional witness of the evolution of Education in the last decades, as well as a regarded channel for the diffusion of the advances in Research and Innovation in the field of Education from a national and international perspective. *Revista de Educación* is published by the Subdirección General de Atención al Ciudadano, Documentación y Publicaciones, and is at present attached to the Instituto Nacional de Evaluación Educativa de la Dirección General de Evaluación y Cooperación Territorial.

GOBIERNO DE ESPAÑA
MINISTERIO DE EDUCACIÓN Y FORMACIÓN PROFESIONAL

NIPO línea: 847-19-002-9
NIPO ibd: 847-19-001-3
ISSN línea: 1988-592X 0034-8082
ISSN papel: 0034-8082

www.educacionyfp.gob.es/revista-de-educacion