

The Recalibration of Public Social Spending in Spain: Towards a Social Investment State?

El recalibrado del gasto social público en España: ¿hacia un Estado orientado a la inversión social?

David Luque Balbona and Ana M. Guillén

Key words

Spain
 • Social Investment
 • Social Protection
 • Recalibration of Social Spending

Palabras clave

España
 • Inversión social
 • Protección social
 • Recalibrado del gasto social

Abstract

Social investment has become the dominant paradigm guiding the reform of European welfare states. Beyond its normative postulates, social investment is a useful analytical framework for the study of ongoing changes in social policy. From this perspective, this article finds evidence of a turn toward social investment in the welfare budget in Spain that began in the mid-nineties. In addition, it shows that the development of new policies linked to social investment have not occurred at the expense of traditional social protection policies. Lastly, it is found that the impact of austerity differs for the different dimensions of social policy.

Resumen

La inversión social se ha consolidado como paradigma dominante para orientar la reforma de los estados de bienestar europeos. Más allá de sus postulados normativos, la inversión social se presenta como un marco analítico útil para el estudio de las transformaciones en marcha en la política social. Desde esta perspectiva, en este trabajo se detecta un giro inversor del presupuesto de bienestar en España desde mediados de los noventa. Además, se muestra que el desarrollo de las nuevas políticas orientadas a la inversión social no se ha hecho a expensas de las políticas tradicionales de protección social. Finalmente, se contrasta el distinto impacto que ha tenido la austeridad en las distintas dimensiones de la política social.

Citation

Luque Balbona, David and Guillén, Ana M. (2021). "The Recalibration of Public Social Spending in Spain: Towards a Social Investment State?". *Revista Española de Investigaciones Sociológicas*, 175: 85-104. (<http://dx.doi.org/10.5477/cis/reis.175.85>)

David Luque Balbona: Universidad de Oviedo | luquedavid@uniovi.es

Ana M. Guillén: Universidad de Oviedo | aguillen@uniovi.es

INTRODUCTION

Social investment has become the dominant paradigm guiding the reform of European welfare states in their adaptation to the emergence of New Social Risks (NSR) associated with post-industrialism (Bonoli, 2005; Taylor-Gooby, 2004). The starting point of the debate over strategies for modernising welfare states can be found in the pioneering works of Giddens (1998) and Esping-Andersen (1999); however, it is only in the last ten years that we find scientific research in this area increasing, particularly after the institutionalisation of the term with the launching of the Social Investment Package for Growth and Social Cohesion in 2013 on the part of the European Commission.

The concept of social investment emerged not only with the ambition to modernise the welfare state, but also to ensure its sustainability (Hemerijck, 2011). Therefore, this perspective on social policy may be of particular importance in overcoming current (and future) challenges to Spanish social policy. However, literature on the matter remains very limited in Spain (Pino, 2014; León and Pavolini, 2014; Rodríguez Cabrero, 2015, 2018; Rodríguez Cabrero, 2015; León *et al.*, 2019; Rodríguez Cabrero, 2018; Guillén and Luque Balbona, 2019).

The main objective of this study is to examine the degree to which a social investment perspective has been implemented in Spain by analysing the evolution of spending in different dimensions of social policy. In addition, we also look at an issue that is often ignored in research: the emergence of NSR does not necessarily imply that Old Social Risks (OSR), belonging to the industrial era, are now less important (Crouch and Keune, 2012; Huber and Stephens, 2007). Thus, we also consider whether the development of new policies oriented toward social investment have been at the expense of traditional social protection poli-

cies. Lastly, we look at whether the Great Recession has led to a shift in the recalibration of the Spanish welfare state.

To achieve these aims, this study is structured in the following form. After this introductory section, the second section presents the notion of social investment as the paradigm guiding social policy reform, develops the analytical framework of the research and describes four hypothetical scenarios for recalibrating social spending. In the third section we analyse the process of recalibrating social spending in Spain over the last two decades, examining whether a shift toward social investment in the social welfare budget has occurred. The fourth section looks at the evolution of spending in social policies that most closely follow a social investment logic. Following, we discuss the obstacles to social investment in the immediate future. The study ends with our conclusions.

SOCIAL INVESTMENT AS PARADIGM AND ANALYTICAL FRAMEWORK

Although the launching of the Social Investment Package for Growth and Social Cohesion by the European Commission in 2013 led to the definitive institutionalisation of the term, the idea of social investment is not that recent (Morel, Palier and Palme, 2012). Its logic already formed part of the Lisbon Agenda, launched in 2000, which had the aim of converting the EU into the most competitive and dynamic knowledge-based economy and capable of achieving sustainable growth, more and better jobs and greater social cohesion (Hemerijck, 2015).

In academic debates, the notion of social investment was initially proposed by Giddens (1998) and Esping-Andersen (1999) in searching for strategies to counteract the NSR emerging from post-industrialism, but

with an important nuance between both versions. In Giddens' *The Third Way*, the (new) social investment state¹ will come to substitute the traditional welfare state, while for Esping-Andersen it will complement it. Therefore, there are two ways of conceiving social investment: with or against social protection; although it seems that "fans of this second solution are more and more numerous" (Barbier, 2017: 51).

In brief, the new welfare state logic that supports the social investment paradigm is that of "preparing" rather than "repairing". The main idea underlying this perspective is that "social policy should no longer focus on 'passively' protecting people *from* the perils of the market by means of cash benefits, but rather prepare or 'empower' people in order to maximally integrate them *into* the market" (Cantillon and Lancker, 2013: 553). Thus, the center of gravity for social policy shifts toward the maximisation of employability and employment (Kersbergen and Hemerijck, 2012). In short, this new paradigm implies a change "from protective welfare states with emphasis on *ex post* remedies to productive welfare states with focus on *ex ante* prevention" (Kuitto, 2016: 445).

From a broad perspective, we can differentiate three interdependent and complementary functions of welfare under a social investment paradigm: 1) alleviating the "flow" of labour market and life-cycle transitions; 2) elevating the quality of the stock of human capital; and 3) maintaining a safety net through minimum incomes as social protection and economic stabilisation "buffers" (Hemerijck, 2014). The function of flow is aimed at a more efficient use and allocation of labour resources over the

life span to achieve high levels of labour market participation. The stock function is aimed at maintaining and improving human capital to foster future productivity. Lastly, the buffering function has two ends: providing adequate social protection and economic stabilisation (Hemerijck, 2015). However, this latter function is often ignored. For the correct development of social investment, the three functions must be integrated. Thus, adequate protection through minimum incomes is a prior critical condition for an effective social investment strategy (Vandenbroucke, Hemerijck and Palier, 2011).

Social investment is not free from criticisms. Regarding the distributive effects of the transformation of the traditional welfare state to a social investment state, there are various authors that suggest it has limited potential for providing greater social cohesion and reducing poverty (Cantillon, 2011; Vandenbroucke and Vleminckx, 2011; Taylor-Gooby, Gummy and Otto, 2015; Vliet and Wang, 2015). Behind these criticisms we find the hypothesis of "resource competition" between different social programmes (Cantillon, 2011). According to this hypothesis, in a period of permanent austerity, the accent placed on new social investment policies shifts some of available resources from (old) protection programmes to (new) active investment programmes (Cantillon, 2011; Vandenbroucke and Vleminckx, 2011).

Cantillon and Lancker (2013) highlight three failures of the social investment perspective to achieve inclusive social progress. First, the corner stone of social investment, achieving social integration through employment, generates a major problem in terms of excluding individuals that cannot participate in the labour market. Secondly, from a social investment perspective, individuals are increasingly responsible for their own welfare. Individuals

¹ In this text we use the term "social investment state"; however, some authors, with the aim of stressing its complementarity with the traditional model, consider it better to use the term "active welfare state" as an "overarching" term for this new paradigm or theoretical approach to social welfare.

are guided toward “appropriate behaviour” by increasing the conditionality of certain social programmes, above all in the labour market and in regards to social assistance benefits. Lastly, some authors point to the existence of a “Matthew effect”, which occurs in some of the leading policies of social investment: “When policy measures are grafted on an underlying logic of (previous) employment, then government investment benefits those already better-off” (Cantillon and Lancker, 2013: 561).

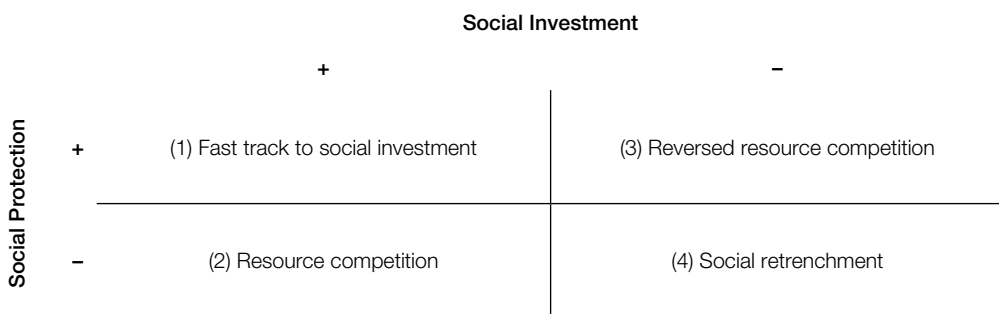
In addition, Jenson (2009), among other authors, argues that the investment perspective gives little attention to gender issues. Fostering women’s economic activity and the development of reconciliation measures are instrumentalised, first, with deomographic ends, and secondly, as a solution to problems generated by the aging of the population. Thus, the aim of fostering equality of opportunity between men and women is not a priority.

Beyond its function as a paradigm guiding social policy reforms, the social investment perspective can be a useful ana-

lytical framework for studying changes in social policy, as it can be used to distinguish forms of social spending that can be classified as “investment” from others that cannot (Nolan, 2013). Thus, we can compare the composition of social spending between countries, its evolution over time and the effectiveness of different forms of spending. In this study we analyse the evolution of specific items of public spending in order to evaluate the transformations that have taken place in social policy (Deken, 2017).

The investment metaphor itself makes the analysis of spending data a good choice for identifying the extent to which governments have adopted a social investment agenda (Deken, 2014; Nolan, 2017). In addition, in an “age of permanent strain” (Pavolinie *et al.*, 2015), a focus on spending is also useful in identifying potential trade-offs between the different dimensions of social policy. In this way, following Ronchi (2018), we find different scenarios or pathways in the recalibration of social spending (Figure 1).

FIGURE 1. Four scenarios for the recalibration of social spending



Source: Adapted from Ronchi (2018: 7).

a) In the upper left quadrant (scenario 1), resources for both dimensions of welfare (protection and investment) increase. In this way, scenario 1 represents a “fast

track to social investment”: a kind of win-win solution that fits well with the perspective defended by Esping-Andersen, among others.

- b) The lower left quadrant (scenario 2) corresponds to the assumption defended by the dominant current in social investment. In this scenario resources are taken from social protection and employed in the investment dimension. This is the “resource competition” hypothesis previously mentioned (Cantillon, 2011; Vandenbroucke and Vleminkx, 2011).
- c) The opposite happens in the upper right quadrant (scenario 3), “reversed competition for resources” (Streeck and Mertens, 2011): resources are taken from the investment dimension and dedicated to the social protection dimension. In this scenario, due to budgetary restrictions and the inertia of traditional (passive) programmes, there is no room for manoeuvre to develop social investment. This scenario is most likely in crisis situations: spending on investment (more discretionary) is more strongly impacted than (obligatory) spending on protection (Breunig and Busemeyer, 2012).
- d) Lastly, the lower right quadrant (scenario 4) refers to “social cuts”: resources dedicated to both investment and protection are reduced. This indicates the arrival of a (genuine) era of retrenchment of social policy.

This study uses public social expenditure data to consider which of the different hypothetical scenarios of recalibration has occurred in Spain. However, prior to this we must make a series of conceptual and methodological clarifications because of the ambiguity of the concept of social investment. First, we look at the conceptual difference between “investment” and others forms of social spending so that, subsequently, we can consider its empirical application.

Following Nolan (2013, 2017), distinguishing between social investment and oth-

ers forms of social spending is more problematic than might initially be thought. First, what is understood by the term investment must be clarified. In economic theory, an investment is spending on goods that are not consumed but that are used for future production, the most obvious examples being the facilities and machinery that workers use to produce goods and services. Therefore, an investment implies postponing consumption to improve productive capacity.

The distinction between investment and consumption is less clear when it is applied to the productive capacity of workers, as “[it] is very difficult to think of a form of social spending that is purely investment, without a substantial element of current consumption”. This is even true when considering education; as the main form of investment in human capital, we cannot negate that “education also clearly has consumption benefits to the individual, in terms of enjoyment of and fulfilment from the educational process itself”. It is also difficult to think of social expenses that are purely consumption.

More broadly, at least some spending by the individual on food, clothing, and so on, counted unambiguously as consumption in national accounting terms, clearly also has a potential return in terms of worker productivity (Nolan, 2013: 464).

To not open this conceptual Pandora’s box, it is best to take investment and consumption as points of reference on a continuum, in which the underlying logic of a concrete type of policy is closer to one extreme (investment) or the other (consumption), and determine their classification in that manner. In other words, we can classify policies in function of the extent they contribute to the type of return that social investment seeks to generate: increasing the size of the labour force and its productivity (Deken, 2014).

How do we approach this empirically? Various studies have aggregated social

expenditure items in different ways to fit them to an analysis from a social investment perspective (among others, Häusermann, 2012; Hemerijck, 2013; Kuitto, 2016; Kvist, 2013; Ronchi, 2018; Vandembroucke and Vleminckx, 2011). Each aggregation presents its own advantages and limitations; see Deken (2014) for a detailed analysis in this respect. However, it is possible to extract a minimum common denominator from the different proposed classifications, although using different labels. Thus, based on the underlying logic of each type of policy and using the terminology employed by Kuitto (2016), certain social spending items have generally been considered as “social investment” as they favour inclusion *in* the labour market (childcare and elderly care policies, active employment policies and education) while others, in contrast, have generally been categorised as “compensatory spending” because they protect persons at risk *from* the market (pensions, contributory benefits for unemployment and monetary transfers to families with children). However, there is a series of policies whose classification in one group or the other is more contingent (for example, parental leaves and health care spending), so that we deliberately exclude them from our analysis. This is due to the “dual purpose” of such policies (Deken, 2014). Thus, for example, spending on healthcare can have an *ex post* compensatory function or an *ex ante* preventative (investive) one. As something new, in this study we incorporate the buffer function in our analysis of social investment through spending on minimum insertion incomes.

To resolve these operational problems in a pragmatic manner, other studies have chosen more generic conceptualisations. Thus, for example, Huber and Stephens (2007) suggest that the debate over the reform of social policy has been, in great part, carried out over terms of an “old”

versus “new” welfare state. The “old” welfare state is conceptualised as centred on transfers, aimed at covering risks connected to the loss of income due to aging, unemployment, illness or disability. The “new” welfare state is conceptualised as centred on services aimed at increasing individuals’ capacities for obtaining income through education, continual training and the requalification and socialisation of care to facilitate balancing work and family life. More recently, Ronchi (2018) also employed this heuristic dichotomy in considering cash benefits as “social protection” and in-kind benefits as “social investment”. This latter option serves us, in what follows, to make an initial approximation of the recalibration of social spending in Spain.

RECALIBRATION OF PUBLIC SPENDING ON SOCIAL POLICIES

In this section we analyse the evolution of public spending on “social protection” and “social investment” policies since the mid-nineties² to obtain a broad overview of the level of development of social investment in Spain. The objective is two-fold: first, we examine if there was some type of recalibration of spending between both dimensions of social policy in the period of economic expansion prior to the crisis and, secondly, we analyse what happened after the sudden change in the economic cycle that took place in 2008. Lastly, we also place the Spanish case in a European con-

² The period of analysis is determined by the availability of time series data on public expenditures in the different areas of social policy necessary to apply the analytical framework we have described in the previous section. However, it is important to emphasise that the recalibration of social spending in Spain began prior to the period under analysis, concretely in the period from 1986 to 1992, during which the welfare state was consolidated in Spain (see among others, Rodríguez Cabrero, 2011).

text. However, prior to this, it is necessary to provide a brief characterisation of the Spanish welfare state to understry its main structural determinants.

According to Ferrera (1996), who grouped Europe's Mediterranean countries together in function of a series of cultural traits, policy dynamics and similar institutional aspects, the Spanish welfare state forms part of a fourth welfare model in Europe. In short, the original conceptual model of types of welfare states was expanded (Esping-Andersen, 1990).

Following González Begega and Luque Balbona, (2016: 263), the countries of southern Europe are characterised by a hybrid type of welfare system, constructed based on the adaptation of institutional pieces and policy logics from other models.

An initial corporatist element, on which the pension system rested [...]. A second social democratic component introduced over the eighties in the healthcare and education systems, although to different degrees of development depending on the country, and a third, liberal or limited social assistance approach.

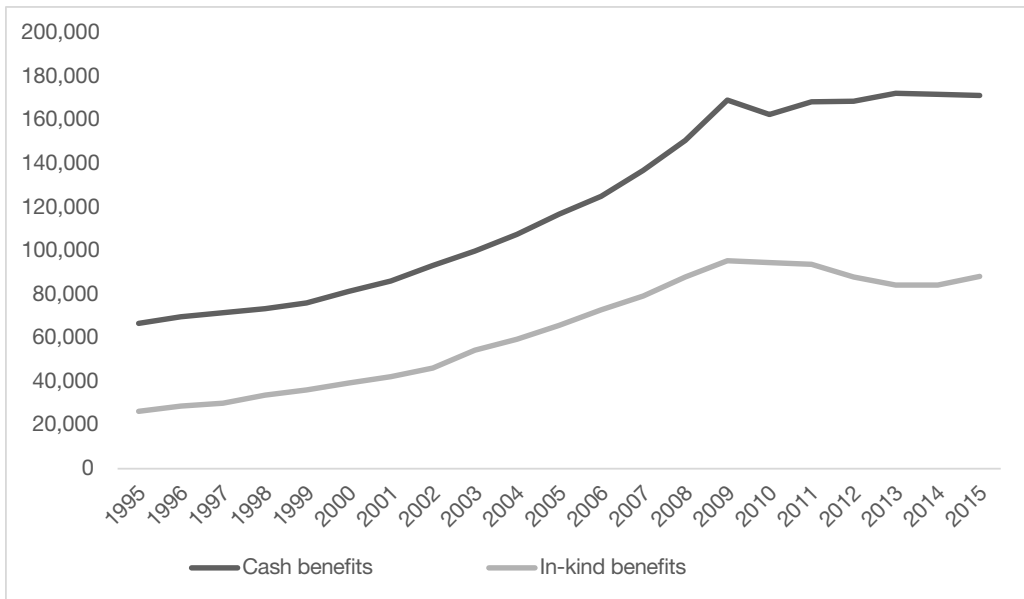
To this, we can add other important distinctive characteristics regarding what interests us here: a "preferential orientation of the social protection system towards pensions, which led to the withdrawal of resources from other areas such as family policies, the underdevelopment of public care services".

Regarding the evolution of public social spending in Spain, as indicated in the previous section, the "old" welfare state is characterised as more oriented toward transfers, while the "new" social investment state is more oriented toward services (Huber and Stephens, 2007). Thus, in a first approximation, in Graph 1 we see the evolution of the distribution of public spending on social protection between cash transfers

(closer to the protective dimension) and in-kind benefits (closer to an investment orientation) since the middle of the nineties.

In the Graph 1 we can see that up until 2009 public spending was growing for both types of benefits, therefore, we find a "fast track" scenario toward social investment. In this period, strong economic growth and healthy public finances led to a wide margin for manoeuvre for expanding spending in social investment without a need for withdrawing resources from other areas of social policy. Thus, in 2009 we find the highest level of social spending on in-kind benefits for the series. This "fast track" path toward social investment shows that, for the Spanish case, for an investment turn to occur it does not necessarily have to be at the expense of (old) compensatory policies.

The international crisis that began in 2008 shifted the country out of the fast track. In Spain, the Great Recession began in the third quarter of 2008 and, technically, ended in the second quarter of 2013. In the middle of that period, in 2010, there was a brief and weak expansion. The recession in Spain was among the most severe, along with Greece, Ireland and Portugal (Bentolila, 2015). In addition, the crisis had a much greater impact on employment than on GDP in Spain, an exception when we look at the international panorama (Garicano, 2012). The overall outcome was the loss of almost 3.3 million jobs between 2008 and 2012 (equivalent to 15.5% of the jobs that had existed at the beginning of the crisis). Some authors connect this disproportionate destruction of employment to the "type of productive specialisation consolidated in the last expansive phase of the economic cycle" (Rocha, 2012: 68). For example, as a consequence of the bursting of the national real estate bubble, the construction sector alone lost 1.4 million jobs.

GRAPH 1. Evolution of public social spending by type of benefit, millions of euros

Source: OECD. Stat (retrieved March 6, 2019).

The crisis also had a strong impact on public finances. Starting from a solid fiscal position (public debt as 35.8% of GDP and a fiscal surplus equivalent to 1.7% of GDP in 2007), during the crisis incomes plummeted while expenses increased as automatic stabilizers went into effect (Laparra and Pérez Eransus, 2012). In the evolution of public spending we can identify two clearly differentiated stages: after an initial expansive stage (2007-2008), a stage of fiscal consolidation (2010-2014) began that strengthened with the intensification of the sovereign debt crisis in the summer of 2011 and the change in government that took place in December 2011 (Conde-Ruiz *et al.*, 2016).

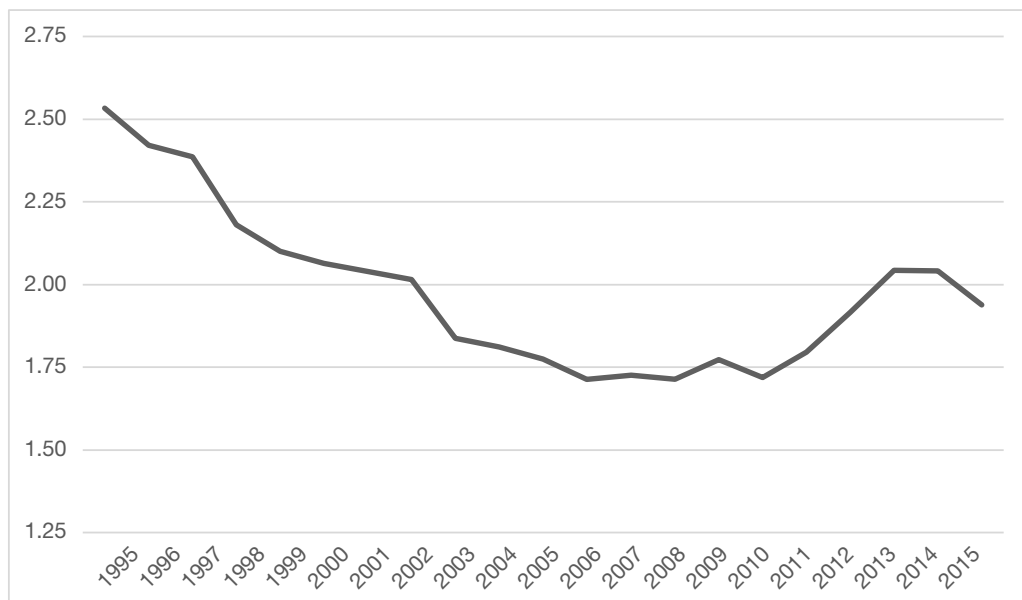
In this context, a dramatic change in the prior trend in public spending on social policy occurred. On the one hand, after a one-off reduction in 2010, linked to the initial adjustment measures announced by the second Zapatero government (PSOE), in May spending on cash benefits slowed its rate of increase since the beginning of the

crisis. On the other hand, in-kind spending, after a slight decrease in the first years of the crisis, declined drastically starting with the beginning of the hard adjustment policies on public spending in 2012, now with Mariano Rajoy (Partido Popular) in the government. As a result, a context of strong austerity regarding public finances led to a scenario of “reversed competition for public resources”; in-kind spending declined, while spending levels on direct cash benefits remained stable as automatic stabilisers came into effect (mainly retirement pensions and unemployment benefits). Thus, the Great Recession introduced a new layer of complexity in the recalibration of social policy (Leoni, 2016). Lastly, with the return to strong economic growth in 2014 and the relaxing of austerity policies, we see an increase in in-kind spending in 2015, while cash transfers remained stable. What this seems to indicate is that the country has not entered into a period of permanent austerity, at least, in terms of social investment spending.

In short, in this initial approach to the recalibration of public social spending we find that the impact of the process of the consolidation of public finances varies in function of the type of social spending. In-kind spending (more connected to social investment) was seriously affected by aus-

terity measures, while spending on monetary benefits (more connected to social protection) only stopped increasing. The return to economic growth, in a consolidated manner since 2014, seems to suggest a recovery in spending on social investment.

GRAPH 2. *Evolution of the ratio spending on cash benefits/spending on in-kind benefits*



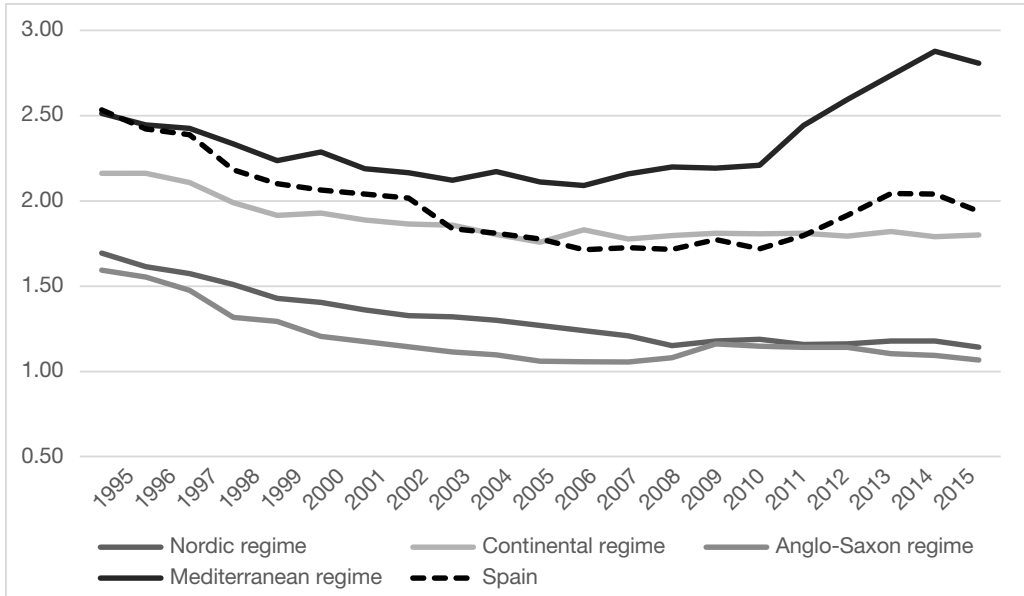
Source: OECD. Stat (retrieved March 6, 2019).

Graph 2 shows the evolution of the ratio (coefficient) between spending on cash benefits and in-kind benefits. From this, we can see that during the first ten years of our period of analysis, the investment side gained in importance with respect to the social protection side: the ratio gradually declined until 2006 (from 2.53 to 1.71). During the period from 2007 through 2010 the process of the recalibration of social spending was suspended: the ratio stabilised, which implies that both types of spending grew in approximately the same proportion. Thereafter, the ratio grew until it reached a level similar to that found for the year 2001 (2.04). As we can see, in this period, spend-

ing on cash benefits remained stable while spending on in-kind benefits declined. In 2015, the last year for which data is available, the ratio resumed its pre-crisis downward trend, spending on in-kind benefits grew and cash benefits remained stable.

To close this section, we look at the Spanish case in its European context. Graph 3 shows the evolution of the ratio between spending on cash benefits and in-kind benefits for the EU15 countries grouped according to their different welfare regimes as identified by Esping-Andersen (1990) and subsequently broadened by Ferrera (1996) by defining the Mediterranean countries as a differentiated group.

GRAPH 3. Evolution of the ratio spending on cash benefits/spending on in-kind benefits for EU15 countries grouped by welfare regimes



Note: Nordic regime: average for Denmark, Finland and Sweden; Continental regime: average for Austria, Germany, Belgium, France, Luxembourg and the Netherlands; Anglo-Saxon regime: average for the United Kingdom; Mediterranean regime: average for Spain, Italy, Greece and Portugal.

Source: OECD. Stat (retrieved March 15, 2019).

Broadly speaking, the evolution of the ratio (protection/investment) in the different regimes confirms previous results from both other researches based on spending data (Kuitto, 2016; Ronchi, 2018) and studies of a more qualitative nature (Bouget *et al.*, 2015). The Nordic and Anglo-Saxon regimes present a more balanced division between the protective and investment dimensions in their welfare budgets. The continental regime is in a middle ground between the former and the Mediterranean countries, although it is closer to the latter, at least until the beginning of the Great Recession. The Mediterranean regime leans more toward the protective dimension; however, Spain continues to follow a path toward recalibration that separates it from this group, as it more and more closely re-

sembles the continental welfare regime. The distancing of Spain from its traditional (Mediterranean) grouping confirms what has been found in other studies (Guillén and León, 2011; Pavolini *et al.*, 2015; Guillén, González Begega and Luque Balbona, 2016). Lastly, the Great Recession only had a drastic impact on the slow process of the recalibration of social spending in the Mediterranean regime; with austerity generating a scenario of “reversed competition for resources”, the progress made during the previous decade disappeared.

In the following section we analyse spending in Spain on items most closely related to the logic of social investment and evaluate the degree of development of its different functions.

EVOLUTION OF PUBLIC SPENDING IN REGARD TO DIFFERENT SOCIAL INVESTMENT FUNCTIONS

In this section we analyse the evolution of public spending on policies most directly connected to social investment. The different spending items are grouped around the three social investment functions (stock, flow and buffers) introduced in the second section of this article.

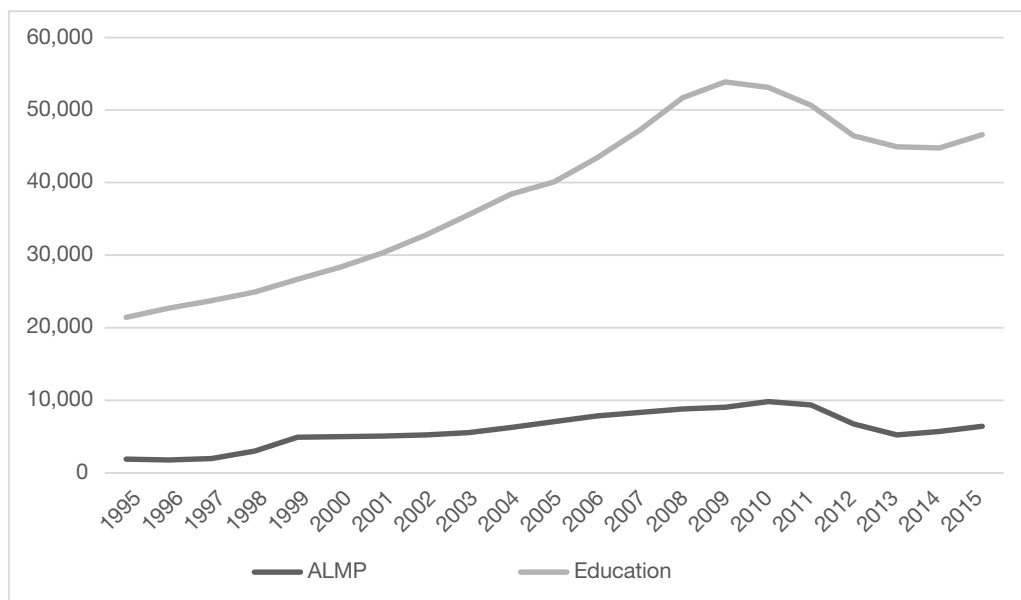
The stock function: the development of human capital

The stock function is connected to future productivity and is aimed at improving, updating and maintaining human capital (Hemerijck, 2015). Education and active labour market policies (ALMP) are most closely connected to this social investment function (Graph 4).

From a social investment perspective, education is a prerequisite for achieving success in the labour market and a foundation for well-being as well (Leoni, 2016). Education can be considered a “pure” form of investment (in human capital) given that it generates returns in the medium and long term in the form of higher rates of employment and better wages.

In Graph 4 we see an important increase in public spending on education up to 2009, with an average annual increase of 7.4% since 2000. In 2014 we find spending of 53.9 billion euros, 90.0% more than in 2000 in current terms. However, with the change in the economic cycle, there was a drastic reduction in public spending on education, primarily concentrated in the 2012-2013 biennium, reaching a low of 44.8 billion euros in 2014 (an accumulated decline of -16.9%). Finally, in 2015 with the relaxing of austerity policies, spending on education increased slightly.

GRAPH 4. Evolution of public spending on the stock function, millions of euros



Note: Spending on all educational levels ISCED 2011 excluding early childhood education (0 to 3 years of age).

Source: Education, Statistics on Public Spending on Education, Ministry of Education (various years). ALMP, OECD. Stat (retrieved March 15, 2019).

For their part, ALMP are also a clear component of social investment strategies. Instead of providing a replacement income to individuals who lose their employment, active policies are based on making the period of unemployment as short as possible, proactively helping those who are unemployed reincorporate into the labour market. In addition, ALMP try to avoid the scarring effect of unemployment and they also generate overall returns to society, reducing spending on benefits and subsidies by fostering reinsertion in the labour market (Bonoli, 2012).

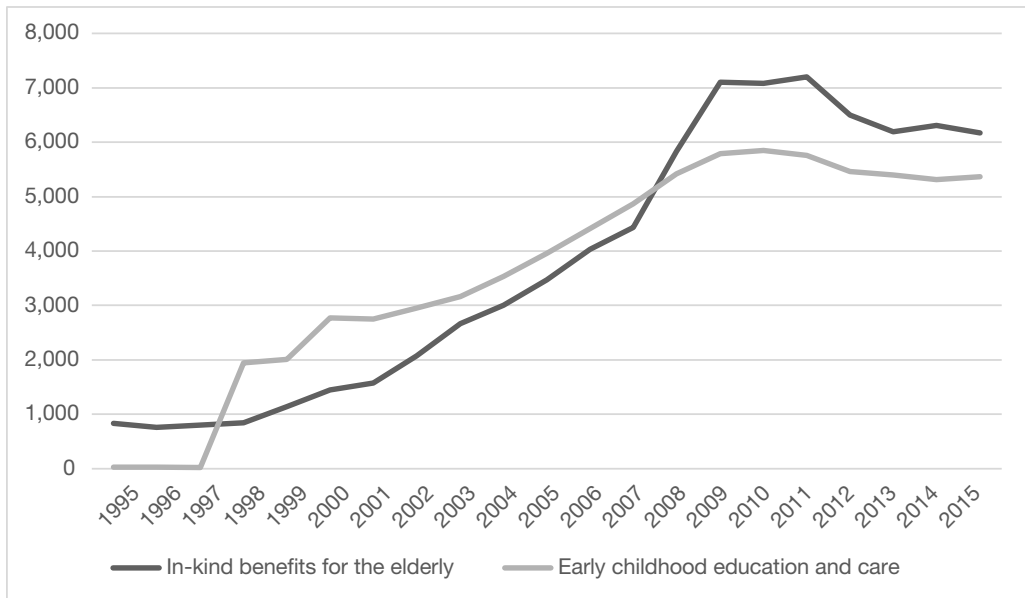
As can be seen in Graph 4 spending on ALMP grew significantly from the mid-nineties until the beginning of the crisis, following the activation principles introduced by the European Employment Strategy (Moreno and Serrano, 2011). With the eruption of the crisis, spending on ALMP

suffered from an evident “competition over resources”, as those required for passive policies increased dramatically with the application of austerity measures. With the consolidation of economic growth in 2014, resources aimed at ALMP resumed their upward growth.

The flow function: care related policies

The flow function is aimed at making the most efficient use of available employment resources and achieving high levels of labour market participation, mainly through facilitating the transition from economic inactivity (Hemerijck, 2015). Thus, the evolution of public spending on *early childhood education and care* and *in-kind benefits for the elderly* are the policies that most closely indicate social investment in this function.

GRAPH 5. Evolution of public spending on the flow function, millions of euros



Source: OECD. Stat (retrieved March 15, 2019).

As can be seen in Graph 5, Spain had been “fertile ground” in the development of care policies (León and Pavolini, 2014) until the arrival of the Great Recession. Spending on early childhood education and care were at minimal levels at the beginning of the nineties but grew significantly and constantly during the years of rapid economic growth and the massive incorporation of women into the labour market (from the mid-nineties until the Great Recession). However, with the initiation of a harsh adjustment to social spending in 2012, the budgetary commitments in these areas declined, although not as abruptly as in other areas of social policy.

Spending on elderly care followed a similar trajectory, starting from low levels in the mid-nineties, spending grew progressively until 2007. In fact, a brief golden age in long-term care began (2007–2011) with the passage of the so-called Dependency Law at the end of 2006, with an accumulated increase in spending of 75.5% with respect to 2006. During the most intense period of austerity (the 2012–2013 biennial), spending declined by 14.0% from its maximum level in 2011. Since then spending in this area has remained relatively stable.

The buffer function: minimum income systems

In Spain, Regional Minimum Incomes programmes (RMIs) are the policies that most closely exercise a buffer function (maintaining strong universal minimum income safety nets). The RMI system is configured as the only protection mechanism for the potentially active population with no previous contributory link and as a safety net for those who have exhausted all other benefits (Arriba, 2014). RMIs are structured as a decentralised system that “operates in different ways depending on the region, leading to very distinct systems, not subject to any process

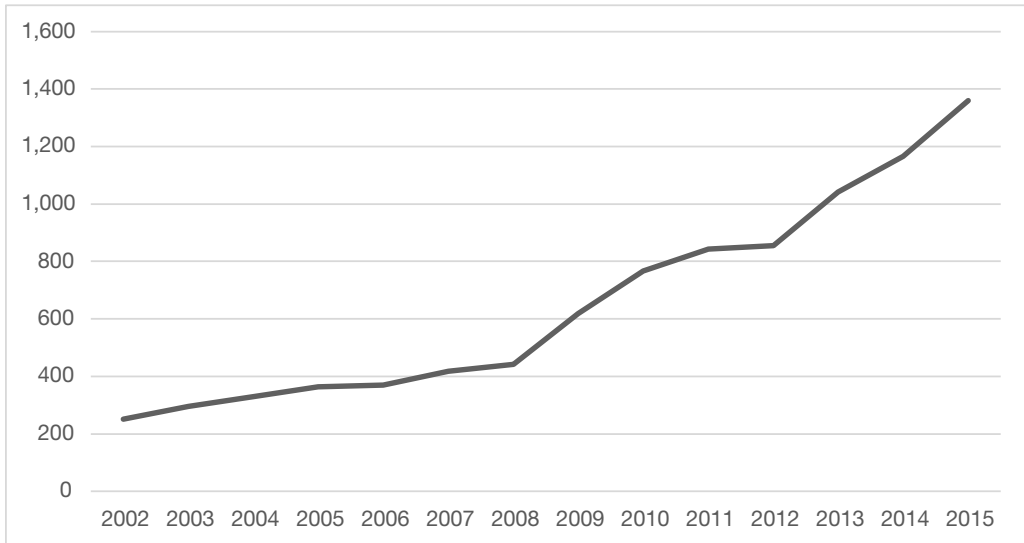
or mechanism of coordination”. In addition, they have a supplementary and complementary dimension, reducing the gaps in protection left uncovered by the central state system (Sanzo, 2018: 43). In Graph 6 we see the evolution in spending on RMIs.

In Graph 6 we can see how, with the crisis, RMIs were no longer the residual systems they had been (Sanzo, 2018). The growth in spending was continuous and particularly intense after the crisis began, increasing from slightly more than 200 million euros in 2002 to almost 1,400 million euros in 2015. As a result, the crisis led to the unexpected institutionalisation of the minimum income system in Spain (Natili, 2016). Thus, the buffer function is the only social investment function that was strengthened during the Great Recession. However, due to the low level of actual protection it provides, the system of guaranteed income is more aimed at avoiding severe poverty than aiding individuals to train for and find employment (Rodríguez Cabrero, 2018).

To summarise our findings in this section, until the Great Recession, Spain was fertile territory for social investment. The three functions (stock, flow and buffer) had all undergone significant development. With the arrival of the crisis, budgetary restrictions led to a tradeoff between the investment and protective dimensions of social policy. As a result, only the buffer function was consolidated during the crisis, while the stock and flow functions suffered from competition for resources.

DISCUSSION: BRAKES ON SOCIAL INVESTMENT

After analysing the evolution of public social spending, we look at a series of factors that can act as a brake on a hypothetical new expansion of social investment in Spain.

GRAPH 6. Evolution of public spending on the buffer function, millions of euros

Source: Report on minimum insertion income (various years). Spanish Ministry of Health, Social Services and Equality.

First, public finances in Spain are still in a delicate situation despite the current favourable economic context. Spain began 2018 as the only EU country with a public deficit above 3% and, therefore, remained under the yoke of the excessive deficit procedure launched in 2009. The meagre public resources available severely limit the expansion of social investment policies, even more so if we take into account the critical state in which we find the pension system, with a deficit, according to Eurostat, of more than 17,000 million euros in 2017. Debate over the future of retirement pensions, part of the paradigm of (old) policies oriented toward protection, occupies a growing portion of public and media attention and social policy focus. The *de facto* repeal³ of the annual revaluation index introduced in a 2013 reform has, in the short term, intensified the problem of the sustainabil-

ity of the system, independent of other considerations. The entry of new pensioners into the system with higher average benefits will require more resources each year, which are not covered by current social security contributions, making it necessary to take resources away from other areas and/or increase indebtedness. This situation may result in a “permanent reversed competition for resources”, in which increased spending on economic benefits (closer to the logic of protection) will not leave space for the expansion of benefits in services (closer to an investment logic). Debate over the sustainability of the pension system shows that the OSR, far from losing relevancy, continue to be central due to the dramatic process of the aging of the population. In this sense, Buesmeyer and Garritzmann (2017: 871) point out that social investment policies “are generally very popular, but as soon as realistic budget constraints are added, public support drops considerably”. Thus, Neimanns, Buse-

³ The annual update of pension amounts in 2018 and 2019 has not been applied.

meyer and Garritzmann (2018) find, in a sample of eight European countries (including Spain), that only 17% of those surveyed were in agreement with increasing spending on social investment at the expense of social protection.

Secondly, beyond the resurgence in OSR, the underlying problem in financing these new social policies linked to a social investment perspective, is the Spanish state's low capacity for raising revenue. Public revenue as a percentage of GDP in Spain is well below the average for its neighbouring countries. Thus, in 2017, public revenue in Spain was 37.9% of GDP, while the average for the EU15 was 45.4% (44.9% in the EU28). As a result, with spending on social protection amounting to 17.1% of GDP in 2015, the latest year with definitive data, the Spanish welfare state can be described as low cost. In addition, Beblavý and Hájková (2016) show that countries with low levels of public revenue, among them Spain, have a relative preference for "compensating social spending". Social investment policies require large and continuing investments that are very difficult for states with low levels of public revenue to undertake. Therefore, one requisite for a (new) expansion of social investment is an increase in the state's revenue collecting capacity.

Lastly, in addition to the issues of the availability of resources and spending preferences between protection and investment, the question arises, as noted by Kazepov and Ranci (2017) for the Italian case, whether the preconditions exist in Spain for an effective social investment strategy. First, one of the weak contextual factors in the Spanish case is the poor connection between the education system (supply of human capital) and the labour market (demand for human capital). Thus, even before the crisis, Spain had the highest level of over-qualification (25%)

among OECD countries (OECD, 2007). This structural disconnection between the education system and the labour market "exposes human capital investments to the risk of over-education and poor economic returns" (Kazepov and Ranci, 2017: 101). Secondly, and also regarding the labour market, there is an incapacity for new jobs to provide sufficient conditions for social integration. Precarious employment is not a result of the crisis, as it was already well-rooted in Spain prior to the crisis. Structural reforms of the labour market and of the collective bargaining system, approved at the height of the crisis, have strengthened the trend toward greater precarisation of employment for a part of the occupied population, as a result of choosing internal devaluation as the economic policy for exiting the crisis. In short, social investment policies are not capable of generating good quality jobs nor of reducing labour market precariousness without the existence of specific contextual conditions connected to the productive fabric and the institutional regulation of the labour market.

CONCLUSIONS

Returning to the objectives discussed in the introduction, from our analysis of public spending data —the skeleton of social investment, according to Deken (2004)— we find that there has been a social investment turn in the welfare budget in Spain over the last two decades. Spending on policies closer to the postulates of social investment has increased in importance over this period. Another important conclusion is that the increase in spending in the investment dimension of social policy has not been at the expense of the protective dimension. Until the Great Recession, Spain, with strong economic growth and sound public finances, was on a "fast

track” toward social investment. Resources aimed at both the (new) investment side as well as the (traditional) protective side increased, although the former more so in relative terms.

In comparative terms, the recalibration of social spending toward social investment has been more intense in Spain than in the other countries of the Mediterranean welfare model. From a social investment perspective, Spain has deviated from the pre-established path, increasing the internal heterogeneity within this group.

Regarding the impact of fiscal austerity on the process of recalibrating social spending, the findings from comparative analyses are confirmed. Austerity measures have more drastically affected (the more discretionary) spending on social investment than (the more consolidated) spending on social protection. In a period of significant budgetary tensions, the trade-off between both dimensions of social policy becomes more evident. Only the social investment buffer function consolidated during the crisis.

Lastly, the most recent data on public spending on social policies indicates that the process of a recalibration toward social investment has resumed with the return to economic growth. However, the low revenue collecting capacity of the Spanish state and the centrality that pensions still occupy in policy debates suggest a reduced margin for manoeuvre to increase this dimension of social policies in the near future.

To conclude, taking the analytical-conceptual development of social investment and the results presented here as a starting point, several aspects not covered in this article can be explored further. For example, given the highly decentralised character of social policy in Spain, particularly in regard to several of the key poli-

cies from a social investment perspective (ALMP, childcare and elderly care policies and RMI), we could examine the potential existence of regional variations in the establishment of a social investment logic. In addition, we could look at citizen preferences regarding the traditional approach to social policy (repair) versus the social investment approach (prepare) to get an idea of the extent of public support for this strategy for modernising social policies.

BIBLIOGRAPHY

- Arriba González de Durana, Ana (2014). “El papel de la garantía de mínimos frente a la crisis”. *VII Informe sobre exclusión and desarrollo social en España 2014*. Madrid: Fundación FOESSA. Available at: <http://www.foessa2014.es/informe>, access January 12, 2019.
- Barbier, Jean-C. (2017). “Social Investment. With or Against Social Protection?”. In: Hemerijck, A. (ed.). *The Uses of Social Investment*. Oxford: Oxford University Press.
- Beblavý, Miroslav and Hájková, Alžbeta (2016). “Social Investment and State Capacity”. (CEPS Working Document No. 419). Available at: <http://aei.pitt.edu/72908/>, access January 10, 2019.
- Bentolila, Samuel (2015). “De recesiones and expansiones”. *Nada es gratis*. Available at: <https://nadaesgratis.es>, access May 2, 2020.
- Bonoli, Giuliano (2005). “The Politics of the New Social Policies: Providing Coverage against New Social Risks in Mature Welfare States”. *Policy & Politics*, 33(3): 431-449.
- Bonoli, Giuliano (2012). “Active Labour Market Policy and Social Investment: A Changing Relationship”. In: Morel, N.; Palier, B. and Palme, J. (eds.). *Towards a Social Investment Welfare State? Ideas, Policies and Challenges*. Bristol: Policy Press.
- Bouget, Denis; Frazer, Hugh; Marlier, Eric and Sabato, Sebastiano (2015). *Social Investment in Europe: A Study of National Policies*. Luxembourg: European Commission–European Social Policy Network (ESPN).
- Breunig, Christian and Busemeyer, Marius R. (2012). “Fiscal Austerity and the Trade-Off between

- Public Investment and Social Spending". *Journal of European Public Policy*, 19(6): 921-938.
- Busemeyer, Marius R. and Garritzmann, Julian L. (2017). "Public Opinion on Policy and Budgetary Trade-Offs in European Welfare States: Evidence from a New Comparative Survey". *Journal of European Public Policy*, 24(6): 871-889.
- Cantillon, Bea (2011). "The Paradox of the Social Investment State: Growth, Employment and Poverty in the Lisbon Era". *Journal of European Social Policy*, 21(5): 432-449.
- Cantillon, Bea and Lancker, Wim van (2013). "Three Shortcomings of the Social Investment Perspective". *Social Policy & Society*, 12(4): 553-564.
- Conde-Ruiz, José I.; Díaz, Manuel; Marín, Carmen and Rubio-Ramírez, Juan (2016). "Evolución del gasto público por funciones durante la crisis (2007-2014): España vs UE". (Working paper, 9, Fedea).
- Crouch, Colin and Keune, Maarten (2012). "The Governance of Economic Uncertainty". In: Bonoli, G. and Natali, D. (eds.). *The Politics of the New Welfare State*. Oxford: Oxford University Press.
- Deken, Johan de (2014). "Identifying the Skeleton of the Social Investment State: Defining and Measuring Patterns of Social Policy Change on the Basis of Expenditure Data". In: Cantillon, B. and Vandenbroucke, F. (eds.). *Reconciling Work and Poverty Reduction*. Oxford: Oxford University Press.
- Deken, Johan de (2017). "Conceptualizing and Measuring Social Investment". In: Hemerijck, A. (ed.). *The Uses of Social Investment*. Oxford: Oxford University Press.
- Esping-Andersen, Gøsta (1990). *The Three Worlds of Welfare Capitalism*. Princeton, New Jersey: John Wiley & Sons.
- Esping-Andersen, Gøsta (1999). *Social Foundations of Postindustrial Economies*. Oxford: Oxford University Press.
- Ferrera, Maurizio (1996). "The 'Southern Model' of Welfare in Social Europe". *Journal of European Social Policy*, 6(1): 17-37.
- Garicano, Luis (2012). "Nuestra Gran Depresión: cinco años de destrucción de empleo". *Nada es Gratis*. Available at: <https://nadaesgratis.es>, access May 2, 2020.
- Giddens, Anthony (1998). *The Third Way: The Renewal of Social Democracy*. Cambridge: University Press.
- Guillén, Ana M. and León, Margarita (eds.) (2011). *The Spanish Welfare State in European Context*. Farnham: Ashgate.
- Guillén, Ana M. and Luque Balbona, David (2019). "La inversión social en España". *Documento de Trabajo del VIII Informe FOESSA*. Madrid: FOESSA.
- Guillén, Ana M.; González Begega, Sergio and Luque Balbona, David (2016). "Austeridad y ajustes sociales en el Sur de Europa. La fragmentación del modelo de bienestar mediterráneo". *Revista Española de Sociología*, 25(2): 261-272.
- Häusermann, Silja (2012). "The Politics of Old and New Social Policies". In: Bonoli, G. and Natali, D. (eds.). *The Politics of the New Welfare State*. Oxford: Oxford University Press.
- Hemerijck, Anton (2011). "The Social Investment Imperative beyond the Crisis". *Challenge Europe*, 21: 11-19.
- Hemerijck, Anton (2013). *Changing Welfare States*. Oxford: Oxford University Press.
- Hemerijck, Anton (2014). "Social Investment, 'stocks', 'flows' and 'buffers'". *Social Policies*, 1(1): 9-26.
- Hemerijck, Anton (2015). "The Quiet Paradigm Revolution of Social Investment". *Social Policy*, 22(2): 242-256.
- Huber, Evelyne and Stephens, John D. (2007). "Combating Old and New Social Risks". In: Armingeon, K. and Bonoli, G. (eds.). *The Politics of Post-Industrial Welfare States*. New York: Routledge.
- Jenson, Jane (2009). "Lost in Translation: The Social Investment Perspective and Gender Equality". *Social Politics*, 16(4): 446-483.
- Kazepov, Yuri and Ranci, Costanzo (2017). "Is Every Country Fit for Social Investment? Italy as an Adverse Case". *Journal of European Social Policy*, 27(1): 90-104.
- Kersbergen, Kees van and Hemerijck, Anton (2012). "Two Decades of Change in Europe: The Emergence of the Social Investment State". *Journal of Social Policy*, 41(3): 475-492.
- Kuitto, Kato (2016). "From Social Security to Social Investment? Compensating and Social Investment Welfare Policies in a Life-Course Perspective". *Journal of European Social Policy*, 26(5): 442-459.
- Kvist, Jon (2013). "The Post-Crisis European Social Model: Developing or Dismantling Social Invest-

- ments?". *Journal of International and Comparative Social Policy*, 29(1): 91-107.
- Laparra, Miguel and Pérez Eransus, Begoña (2012) (coords.). *Crisis and fracture social en Europa. Causas and efectos en España*. Barcelona: Obra Social "La Caixa".
- León, Margarita and Pavolini, Emmanuele (2014). "Social Investment or Back to Familism: The Impact of the Economic Crisis on Family and Care Policies in Italy and Spain". *South European Society and Politics*, 19(3): 353-369.
- León, Margarita; Ranci, Constanzo; Sabatinelli, Stefania and Ibáñez, Zyab (2019). "Tensions between Quantity and Quality in Social Investment Agendas: Working Conditions of ECEC Teaching Staff in Italy and Spain". *Journal of European Social Policy*, 29(4): 564-576.
- Leoni, Thomas (2016). "Social Investment: A Guiding Principle for Welfare State Adjustment after the Crisis?". *Empirica*, 43(4): 831-858.
- Morel, Nathalie; Palier, Bruno and Palme, Joakim (eds.). (2012). *Towards a Social Investment Welfare State?: Ideas, Policies and Challenges*. Bristol: Policy Press.
- Moreno, Luis and Serrano, Amparo (2011). "Europeanization and Spanish Welfare: The Case of Employment Policy". In: Guillén, A. M. and León, M. (eds.). *The Spanish Welfare State in European Context*. Farnham: Ashgate.
- Natili, Marcello (2016). "The Unexpected Institutionalization of Minimum Income in Spain". *Laboratorio di Politica Comparata e Filosofia Pubblica*. Turin: Centro Einaudi. (Working Paper No. 2).
- Neimanns, Erik; Busemeyer, Marius R. and Garritzmann, Julian L. (2018). "How Popular Are Social Investment Policies Really? Evidence from a Survey Experiment in Eight Western European Countries". *European Sociological Review*, 34(3): 238-253.
- Nolan, Brian (2013). "What use is 'social investment'?". *Journal of European Social Policy*, 23(5): 459-468.
- Nolan, Brian (2017). "Social Investment: The Thin Line between Evidence-Based Research and Political Advocacy". In: Hemerijck, A. (ed.). *The Uses of Social Investment*. Oxford: Oxford University Press.
- OCDE (2007). *International Migration Outlook*. Paris: OCDE.
- Pavolini, Emmanuele; León, Margarita; Guillén, Ana M. and Ascoli, Ugo (2015). "From Austerity to Permanent Strain? The EU and Welfare State Reform in Italy and Spain". *Comparative European Politics*, 13(1): 56-76.
- Pino, Eloisa del (2014). "¿Un Estado orientado a la inversión social? Avances y retrocesos de los Gobiernos europeos en relación con el nuevo paradigma del bienestar". *Gaceta Sindical: Reflexión and Debate*, 22: 141-156.
- Rocha, Fernando (2012). "La crisis económica y sus efectos sobre el empleo en España". *Gaceta Sindical*, 19: 67-90.
- Rodríguez Cabrero, Gregorio (2011). "The Consolidation of the Spanish Welfare State". In: Guillén, A. M. and León, M. (eds.). *The Spanish Welfare State in European Context*. Farnham: Ashgate.
- Rodríguez Cabrero, Gregorio (coord.) (2015). *ESPN Thematic Report on Social Investment – Spain*. Brussels: European Commission. Available at: <http://ec.europa.eu/social/keyDocuments.jsp>, access June 1, 2018.
- Rodríguez Cabrero, Gregorio (2018). "Políticas sociales e inversión social en un horizonte de consolidación fiscal y reestructuración permanente del estado de bienestar". In: Jaraíz Arroyo, G. (coord.). *Bienestar social y políticas públicas: retos para pactar el futuro*. Madrid: La Catarata.
- Ronchi, Stefano (2018). "Which Roads (if any) to Social Investment? The Recalibration of EU Welfare States at the Crisis Crossroads (2000-2014)". *Journal of Social Policy*, 47(3): 459-478.
- Sanzo González, Luis (2018). "La política de garantía de ingresos en España". *Revista de Servicios Sociales*, 65: 41-51.
- Streeck, Wolfgang and Mertens, Daniel (2011). *Fiscal austerity and public investment: Is the possible the enemy of the necessary?* (MPIfG Discussion Paper 11/12).
- Taylor-Gooby, Peter (2004). "New Social Risks and Welfare States: New Parading and New Politics?". In: Taylor-Gooby, P. (ed.). *New Risks, New Welfare. The Transformation of the European Welfare State*. Oxford: Oxford University Press.
- Taylor-Gooby, Peter; Gumy, Julia M. and Otto, Adelaide (2015). "Can 'New Welfare' Address Poverty through More and Better Jobs?". *Journal of Social Policy*, 44(1): 83-104.
- Vandenbroucke, Frank and Vleminckx, Koen (2011). "Disappointing Poverty Trends: Is the Social Investment State to Blame?". *Journal of European Social Policy*, 21(5): 450-471.

Vandenbroucke, Frank; Hemerijck, Anton and Palier, Bruno (2011). "The EU Needs a Social Investment Pact". *OSE Paper Series*. (Opinion paper No. 5).

Vliet, Olaf van and Wang, Chen (2015). "Social Investment and Poverty Reduction: A Comparative Analysis across Fifteen European Countries". *Journal of Social Policy*, 44(3): 611-638.

RECEPTION: March 28, 2019

REVIEW: March 24, 2020

ACCEPTANCE: July 28, 2020

La calidad de las preguntas de encuesta en España: una comparación transnacional

The Quality of Survey Questions in Spain: A Cross-National Comparison

Oriol J. Bosch y Melanie Revilla

Palabras clave

Calidad de los datos
 • Errores de medición
 • Experimento
 MultiRasgo-
 MultiMétodo
 • Investigación
 transnacional
 • Metodología de
 encuestas

Key words

Data Quality
 • Measurement Errors
 • MultiTrait-
 MultiMethod
 Experiment
 • Cross-National
 Research
 • Survey Methodology

Resumen

La mayoría de la investigación social estudia conceptos abstractos (p. ej., actitudes) mediante preguntas de encuestas. No obstante, las encuestas adolecen de errores de medición que afectan a las conclusiones sustantivas. Cuando dichos errores difieren entre países, comparar relaciones estadísticas estandarizadas entre países puede resultar en conclusiones incorrectas. Sin embargo, la calidad de medición de las preguntas de encuestas en España no ha sido investigada de forma comparada. Utilizando un experimento MultiRasgo-MultiMétodo, realizado en la Encuesta Social Europea (ESS), comparamos la calidad de las preguntas en España con la de otros países. En general, la calidad de medición en España es superior a la mayoría de países participantes. Además, si no se tienen en cuenta los errores de medición al comparar España con otros países, las conclusiones sustantivas pueden ser erróneas.

Abstract

Most social research collects data about abstract concepts (e.g., attitudes) using survey questions. However, survey data suffer from measurement errors that affect substantive conclusions. When measurement errors differ across countries, cross-national comparisons of standardized relationships can result in incorrect substantive conclusions. However, no research has analysed the measurement quality of survey questions in Spain in a comparative perspective. Using a Split-Ballot Multitrait-Multimethod experiment conducted in the European Social Survey round 8, we compare the quality of questions in Spain with their quality in other participating countries. The average measurement quality in Spain is higher than the overall average for all ESS countries. In addition, when comparing Spain with other countries, substantive conclusions can be incorrect if differences in the size of measurement errors are not taken into account.

Cómo citar

Bosch, Oriol J. y Revilla, Melanie (2021). «La calidad de las preguntas de encuesta en España: una comparación transnacional». *Revista Española de Investigaciones Sociológicas*, 175: 3-26. (<http://dx.doi.org/10.5477/cis/reis.175.3>)

La versión en inglés de este artículo puede consultarse en <http://reis.cis.es>

Oriol J. Bosch: The London School of Economics and Political Science y Research and Expertise Centre for Survey Methodology (RECSM) - Universitat Pompeu Fabra | o.bosch-jover@lse.ac.uk

Melanie Revilla: Research and Expertise Centre for Survey Methodology (RECSM) - Universitat Pompeu Fabra | melanie.revilla@upf.edu

INTRODUCCIÓN¹

La mayoría de la investigación social requiere de la recopilación de datos sobre conceptos abstractos como actitudes, sentimientos u opiniones. Estos conceptos, que corresponden a representaciones mentales no directamente observables, suelen operacionalizarse mediante indicadores empíricos, siendo las preguntas de encuestas el tipo más común (Sarís y Gallhofer, 2014).

Una operacionalización adecuada de estos conceptos implica diseñar preguntas que maximicen la fuerza de la relación estadística entre el concepto latente que los investigadores quieren medir (por ejemplo, felicidad, F) y los indicadores observados (respuestas a las preguntas, Y). La fuerza de esta relación entre F e Y , cuando está estandarizada, se llama calidad de medición (q^2) y puede calcularse como el producto de la fiabilidad (r^2) y la validez (v^2) (Sarís y Andrews, 1991). La fiabilidad representa la fuerza de la relación entre la respuesta observada (Y) y el valor verdadero (T), es decir, el valor de una pregunta de encuesta con una escala determinada si no se hubieran producido errores aleatorios al responder. La validez representa la fuerza de la relación estadística entre el concepto de interés latente (F) y el valor verdadero (T) de una pregunta determinada. La calidad de medición toma valores de 0 a 1.

Idealmente, la calidad de medición debería ser igual a 1 (la pregunta mide perfectamente el concepto de interés). Sin embargo, en la práctica, los datos de las encuestas adolecen de errores de medición

aleatorios y sistemáticos, que son el complemento de la calidad de medición y, por lo tanto, se pueden calcular como $1 - q^2$.

Alwin (2007) sugiere que el 50% de la varianza (es decir, la dispersión o variabilidad de la distribución) de las variables observadas en las encuestas se debe a errores de medición. Por tanto, existen grandes diferencias entre la variable que los investigadores quieren medir (F) y la que realmente mide la pregunta (Y).

El tamaño de estos errores de medición depende de cómo se diseñan las preguntas de encuesta (por ejemplo, formulación exacta o escalas de respuesta), el idioma y el país donde se administra la encuesta (Liao, Sarís y Zavala-Rojas, 2019), el modo (cara a cara, teléfono, etc.) de recopilación de datos y, para las encuestas *online*, también el tipo de dispositivo utilizado para responder (Bosch *et al.*, 2019). Esto, a su vez, puede tener serias implicaciones para las conclusiones de la investigación. Sarís y Gallhofer (2007) ilustraron este punto utilizando datos de un experimento realizado en la ronda 1 de la Encuesta Social Europea (European Social Survey, ESS) en Gran Bretaña: mientras que la correlación entre la confianza interpersonal y la confianza en el Parlamento medida usando una escala de cuatro puntos era negativa y significativa (-0,15), al usar una escala de 11 puntos la misma correlación era positiva y significativa (0,29). Sin embargo, ambas correlaciones contienen errores de medición. Para saber cuál es la verdadera correlación entre confianza interpersonal y confianza en el Parlamento, es necesario obtener información sobre el tamaño de los errores de medición de las diferentes escalas para corregir estos (Sarís y Gallhofer, 2014). Sin embargo, Sarís y Revilla (2016) encontraron que, para varias revistas importantes de ciencias sociales y *marketing*, solo el 9% de los estudios que utilizaron datos de encuestas corrigió dichos errores.

¹ Agradecimientos: Queremos agradecer al equipo científico central (Core Scientific Team, CST) de la Encuesta Social Europea por su apoyo continuo a esta línea de investigación. Esta investigación ha sido financiada por el ESS ERIC Work Programme 1 de junio 2017 - 31 de mayo 2019.

Al realizar investigaciones transnacionales, los errores de medición pueden afectar la comparabilidad de los resultados entre países. Cuando los errores de medición varían de un país a otro, las comparaciones de relaciones estandarizadas entre países pueden dar lugar a conclusiones sustantivas erróneas (Saris y Revilla, 2016). Según Saris y Gallhofer (2007), las principales características de una pregunta que pueden variar entre países y, en consecuencia, provocar diferencias en la calidad de medición son: 1) las características lingüísticas, 2) los niveles de deseabilidad social y 3) el nivel de centralidad de dicha pregunta. Con respecto a las diferencias lingüísticas, los idiomas tienen distintas estructuras, lo que puede conducir a diferentes niveles de calidad de medición entre países, incluso si se traducen correctamente (Zavala-Rojas, 2016). Igualmente, la deseabilidad social, es decir, la tendencia de los encuestados a responder de una manera que consideran socialmente más aceptable que su respuesta «verdadera» (DeMaio, 1984), muestra diferencias interculturales sistemáticas (Johnson y Vijver, 2003), siendo más alto, en particular, en sociedades colectivistas. Finalmente, los temas de las preguntas pueden tener diferentes niveles de importancia o estar más o menos presentes en el debate público, lo que significa que su centralidad (o prominencia), es decir, el grado en que el tema de cualquier pregunta resuena con el encuestado y la cantidad de información disponible, también puede variar entre países (Couper y Leeuw, 2003).

La investigación realizada hasta ahora, por ende, sugiere que existen diferencias en el tamaño de los errores de medición entre países y que, dichas diferencias, pueden afectar a las comparaciones entre países. Aun con todo, solo unos pocos estudios han explorado las diferencias transnacionales en cuanto al tamaño de los errores de medición.

Este artículo contribuye de varias formas a enriquecer la escasa literatura existente sobre las diferencias en el tamaño de los errores de medición entre países. En primer lugar, nos centramos en comparar España con otros países europeos. Por un lado, en abril de 2019, España era un participante fijo o rotativo de al menos 21 encuestas transnacionales, activas hoy en día, centradas en muestras —de individuos u hogares privados— de la población general (GESIS, 2019). Asimismo, en agosto de 2019, España era el quinto país con más usuarios registrados utilizando datos de la ESS y el sexto en términos de descargas de datos. Además, 77 publicaciones científicas han utilizado datos de la ESS de España hasta agosto de 2019 (ESS, 2019). Así pues, abundante investigación transnacional se realiza utilizando dichos datos.

Por otro lado, existen evidencias de que España puede diferir en términos de la calidad de los datos de encuestas en comparación con otros países europeos. Por ejemplo, las tasas de respuesta (que a menudo se utilizan como un indicador de la calidad de los datos) disminuyeron o se estancaron en la mayoría de los países participantes de las rondas 1 a 7 de la ESS, mientras que aumentaron en España (Beullens *et al.*, 2018). Para otros indicadores comúnmente utilizados para inferir la calidad de los datos, tales como la aquiescencia y el estilo de respuesta extremo, se ha descubierto que son más presentes en países mediterráneos como España que en otros países europeos como Alemania o Gran Bretaña (Herk, Poortinga y Verhallen, 2004). Esto podría estar relacionado con el hecho de que la aquiescencia aumenta cuando los niveles de colectivismo y corrupción son más elevados en un país (Rammstedt, Danner y Bosnjak, 2017), presentando España niveles moderados de colectivismo (Beilmann, Kööts-Ausmees y Realo, 2018; Leung *et al.*, 1992) y de percepción de la

corrupción (Transparency International, 2019). Además, teniendo en cuenta que la deseabilidad social es mayor en las sociedades colectivistas (Johnson y Vijver, 2003), esto podría dar lugar a niveles de errores de medición en España distintos comparado con otros países europeos. En general, pues, se puede esperar que España muestre una calidad de datos diferente a la de otros países europeos.

Sin embargo, muy pocos estudios han analizado la calidad de medición (q^2) de las preguntas de encuesta (como se definió anteriormente) en España en comparación con otros países europeos, con dos notables excepciones:

- 1) Saris *et al.* (2010), utilizando experimentos MultiRasgo-MultiMétodo (*MultiTrait-MultiMethod*, MTMM) de las rondas 2 (2004) y 3 (2006) de la ESS, estimaron la calidad de medición de 12 preguntas sobre cuatro temas: «la distancia social entre médicos y pacientes», «opinión sobre el trabajo», «opinión sobre políticas de inmigración» y «opinión sobre las consecuencias de la inmigración». Los autores descubrieron que, en general, España tiene una calidad de medición superior a la media de la ESS.
- 2) Revilla, Saris y Krosnick (2014), utilizando experimentos MTMM de la ronda 3 de la ESS, estimaron la calidad de 12 preguntas sobre cuatro temas: los mismos temas mencionados anteriormente de «opinión sobre políticas de inmigración» y «opinión sobre consecuencias de la inmigración», así como «sentimientos sobre la vida y las relaciones» y «apertura al futuro». Encontraron una calidad de medición superior en España a la media de los países participantes en dicha ronda de la ESS.

Sin embargo, ambos artículos son muy específicos respecto al tipo de comparaciones que les interesan (respectivamente, es-

calas «de acuerdo-en desacuerdo» [*agree-disagree scales*] versus escalas específicas [*item-specific scales*] y variaciones en el número de categorías de respuesta en escalas «de acuerdo-en desacuerdo»). Además, ninguno de ellos se centra en las diferencias entre países ni en las implicaciones de estas diferencias para la investigación transnacional, específicamente cuando se trata de comparar España con otros países europeos.

En segundo lugar, nos centramos, a diferencia de anteriores estudios, en distintas características de las preguntas (por ejemplo, el nivel de correspondencia entre los números y las etiquetas verbales o el hecho de mostrar las preguntas en las tarjetas que se facilitan a los participantes). De esta manera, podemos brindar información útil para ayudar a diseñar diferentes aspectos de los cuestionarios, sobre los que aún falta evidencia empírica.

En tercer lugar, ilustramos las implicaciones para la investigación sustantiva (en particular transnacional) de no tener en cuenta los errores de medición. Si bien investigaciones anteriores (por ejemplo, Saris y Revilla, 2016) presentaron un método para corregir los errores de medición, las aplicaciones prácticas para la investigación transnacional aún son escasas.

Por último, proporcionamos recomendaciones prácticas a investigadores y profesionales interesados en realizar investigaciones transnacionales utilizando datos de encuestas de España. Estas recomendaciones son útiles tanto para los investigadores que diseñan sus propios cuestionarios como para aquellos que utilizan datos de encuestas existentes, como la ESS. Para ello, utilizamos datos de un experimento MTMM sobre «actitudes hacia la calificación de entrada o exclusión de inmigrantes» que se realizó mediante entrevistas presenciales en 23 países durante la ronda 8 de la ESS (2016-2017).

MÉTODO

El modelo *True Score* MTMM

Para explorar la calidad de medición en España en comparación con otros países europeos, estimamos la calidad de medición utilizando datos de un experimento MTMM. El enfoque MTMM, introducido por primera vez por Campbell y Fiske en 1959, consiste en repetir un conjunto de preguntas que miden conceptos latentes simples correlacionados entre ellos (por ejemplo, opiniones sobre inmigración), llamados rasgos (F_i), utilizando varios métodos (M_j). En 1971, Jöreskog propuso tratar las matrices MTMM como un modelo de Análisis Factorial Confirmatorio (*Confirmatory Factor Analysis*, CFA). En 1984, Andrews sugirió utilizar el enfoque MTMM para evaluar la calidad de medición de preguntas individuales a través de Modelos de Ecuaciones Estructurales (*Structural Equation Modeling*, SEM), utilizando un modelo en el que los efectos de los métodos se suman (*additive method effect model*). En contraste, Browne (1984) y Cudeck (1989) propusieron un modelo en el que esos efectos se multiplican (*multiplicative method effect model*). Corten *et al.* (2002) mostraron que un modelo aditivo dependiente de la escala (*scale-dependent additive model*) funciona mejor que otros cuatro modelos multiplicativos y/o invariantes de escala (*scale-invariant*). Por otro lado, Saris y Aalberts (2003) demostraron que la presencia de efectos de método es una mejor explicación para los términos perturbativos correlacionados en los experimentos MTMM en comparación con otras posibles explicaciones como las respuestas relativas, la aquiescencia o variaciones en las funciones de respuesta. Por lo tanto, en este estudio utilizamos un modelo en el que los efectos de los métodos se suman y son dependientes de la escala (*scale-dependent additive method effects model*). Siguiendo el enfoque de Andrews (1984), considera-

mos que cada método corresponde a una escala de respuesta (por ejemplo, escala de 6 puntos o de 11 puntos) y que los mismos encuestados responden a las mismas preguntas varias veces, utilizando los diferentes métodos. Más precisamente, utilizamos el modelo *True Score* («Valor verdadero») propuesto por Saris y Andrews (1991) que además permite estimar por separado los coeficientes de fiabilidad, validez y método. Esto es una ventaja, ya que a menudo se ven afectados de manera diferente por los cambios en las características de la pregunta.

El modelo *True Score* se puede resumir con el siguiente sistema de ecuaciones:

$$Y_{ij} = r_{ij} T_{ij} + e_{ij} \quad (1)$$

$$T_{ij} = v_{ij} F_i + m_{ij} M_j \quad (2)$$

donde F_i es el rasgo o factor i , M_j es el método j , Y_{ij} es la respuesta observada por el rasgo i y el método j , T_{ij} es el componente sistemático de la respuesta por el rasgo i y método j llamado *true score factor*, r_{ij} es el coeficiente de fiabilidad (cuando se estandariza), v_{ij} es el coeficiente de validez (cuando se estandariza), y e_{ij} es el error aleatorio asociado con Y_{ij} .

La ecuación (1) define cada variable observada (Y_{ij}) como la suma de los asociados *true score* (T_{ij}) y los errores aleatorios (e_{ij}). La ecuación (2) indica que cada *true score* (T_{ij}) es en sí mismo la suma del componente del rasgo (F_i) y el efecto de método usado para medirlo (M_j).

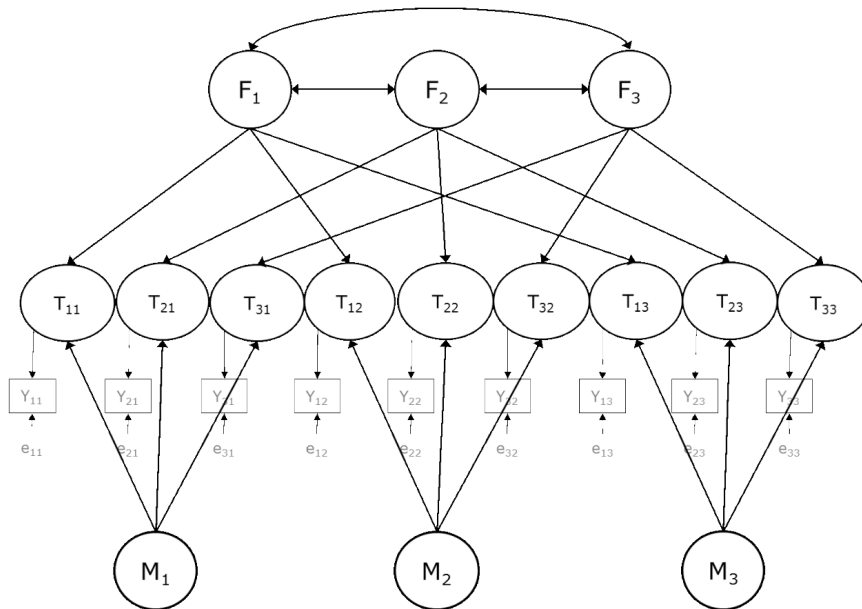
Como punto de partida para este modelo, asumimos que: a) los errores aleatorios no están correlacionados entre sí ni con las variables independientes en las diferentes ecuaciones, b) los rasgos están correlacionados, c) los factores del método no están correlacionados entre ellos ni con los rasgos, y d) el impacto del factor del método sobre los rasgos medidos con una escala común es el mismo. Al testear el

modelo, algunas de las asunciones hechas en este modelo base se pueden relajar si es necesario (ver sección «Análisis y testeo de los MTMM»), hasta que se pueda obtener un modelo final con buen ajuste.

Una vez hecho esto, la calidad total de medición se puede obtener tomando el producto de la fiabilidad y la validez (el cuadrado de sus coeficientes): $q_{ij}^2 = r_{ij}^2 * v_{ij}^2$

En el interés de conseguir un modelo identificado, el modelo MTMM generalmente repite tres rasgos, cada uno medido con tres métodos, resultando en nueve variables observadas. Por tanto, cada encuestado debe responder la misma pregunta tres veces con diferentes escalas. La figura 1 ilustra un modelo *True Score* MTMM para tres rasgos y tres métodos.

FIGURA 1. Modelo True Score MTMM para tres rasgos y tres métodos



Fuente: Elaboración propia.

El enfoque *Split-Ballot* MTMM

Con el fin de reducir la carga cognitiva de los encuestados y los posibles efectos de memoria debido a la repetición de las mismas preguntas a los mismos encuestados (Meurs y Saris, 1990), Saris, Satorra y Coenders (2004) propusieron combinar el enfoque MTMM con un diseño donde los encuestados se asignan al azar a varios grupos (llamado diseño *Split-Ballot*; SB), los cuales reciben un cuestionario ligeramente distinto al otro. Cada grupo obtiene

una combinación de dos métodos para un conjunto dado de tres rasgos, en lugar de obtener tres métodos. Con esto se pueden estimar todos los coeficientes de fiabilidad y validez. El modelo es normalmente identificado en condiciones generales cuando se utiliza un diseño de *Split-Ballot* (Saris, Satorra y Coenders, 2004). Es posible dividir a los encuestados en diferentes números de grupos, incluso con tamaños de muestra desiguales (Revilla, Bosch y Weber, 2019).

Dado que los problemas de no convergencia y valores inválidos ocurren con fre-

cuencia para el diseño de dos grupos (Revilla y Saris, 2013), en la ronda 8, la ESS implementó un diseño de tres grupos. El grupo 1 respondió al método 1 (M_1) en el momento 1 (Sección C) y al método 2 (M_2) en el momento 2 (Sección I). El grupo 2 respondió al M_2 en el momento 1 y al método 3 (M_3) en el momento 2. Finalmente, el grupo 3 respondió al M_3 en el momento 1 y al M_1 en el momento 2. Con este diseño se observan todas las posibles correlaciones entre métodos.

Análisis y testeo de los MTMM

Los coeficientes de fiabilidad y validez se estiman usando CFA (modelo *True Score* presentado anteriormente) y LISREL 8,72 (Jöreskog y Sörbom, 1996). LISREL utiliza algoritmos complejos que minimizan los residuos, teniendo en cuenta todas las restricciones del modelo. El método utilizado para la estimación en cada país es el de Máxima Verosimilitud (*Maximum Likelihood*) para análisis de grupos múltiples (los diferentes grupos son los grupos de *Split-Ballot*). Nos referimos al apéndice A para un ejemplo del código base de LISREL y a Hox y Bechger (1998) para una introducción en profundidad a los Modelos de Ecuaciones Estructurales.

Para probar si hay problemas debidos a especificaciones incorrectas, utilizamos el *software* JRule (Veld, Saris y Satorra, 2008) basado en el procedimiento desarrollado por Saris, Satorra y Veld (2009). JRule tiene la ventaja de tener en cuenta el poder estadístico (es decir, la probabilidad de aceptar una hipótesis nula falsa). También testea las especificaciones incorrectas a nivel de parámetro, es decir, testea si cada parámetro está mal especificado en vez de testear todo el modelo a la vez.

Esto lleva en muchos casos a la introducción de correcciones con respecto a las asunciones del modelo base presentado en las ecuaciones 1 y 2. Principalmente, los

cambios consisten en: 1) permitir efectos desiguales de un método sobre los valores verdaderos correspondientes a los diferentes rasgos, 2) liberar las varianzas de los términos de error entre los grupos de *Split-Ballot* para tener en cuenta el hecho de que los errores aleatorios pueden diferir en los momentos 1 y 2 (por ejemplo, porque los encuestados se cansan con el tiempo), 3) agregar una correlación entre dos factores de método con características de escalas similares, y 4) permitir correlaciones entre varianzas de los términos de error debido a efectos de memoria. Para poder comparar los resultados entre países e idiomas, primero consideramos introducir correcciones similares en los diferentes grupos de países e idiomas. Sin embargo, no es siempre posible. Las correcciones finales del modelo realizadas en cada análisis se resumen en el apéndice B, junto con diferentes indicadores del ajuste del modelo.

Después de realizar los análisis y testear los modelos MTMM, calculamos la calidad de medición para los diferentes rasgos y métodos.

Corrección por errores de medición

Las relaciones estadísticas estandarizadas entre variables observadas, como las correlaciones o los coeficientes de regresión, se ven afectadas por los errores de medición. Por ejemplo, Saris y Revilla (2016), utilizando datos de la ronda 3 de la ESS de Gran Bretaña, encontraron que la correlación de «permitir que más inmigrantes vengan a Gran Bretaña» con la opinión de que los inmigrantes hacen del país un lugar peor para vivir cambió de $-0,27$ (sin corrección) a $-0,61$ (corrección), mientras que la correlación con la opinión de que la inmigración es mala para la economía pasó de $0,13$ (sin corrección) a $0,00$ (corrección).

Para estimar las verdaderas relaciones (es decir, las relaciones entre los concep-

tos de interés), es necesario corregir por los errores de medición. En el marco de las comparaciones entre países, asimismo, un requisito para comparar las relaciones estadísticas estandarizadas usando variables observadas es tener niveles similares de calidad de medición en dichos países. Dicho de otra manera, si el tamaño de los errores de medición difiere entre países, no se debería realizar comparaciones directas de relaciones estadísticas estandarizadas sin corregir primero por los errores de medición.

Hay diferentes formas de corregir por los errores de medición (véanse DeCastellarnau y Saris, 2014; Saris y Gallhofer, 2014). La corrección de dichos errores de medición se puede realizar para diferentes tipos de análisis, incluidas correlaciones, regresiones lineales simples, SEM, etc. Este artículo se centra en una ilustración que compara la correlación entre dos conceptos simples, cada uno medido por una sola pregunta². En el caso que ilustramos, imaginemos que unos investigadores observan la correlación entre las respuestas a dos preguntas individuales $\rho(Y_1, Y_2)$, pero están interesados en la correlación entre los conceptos latentes detrás de cada una de las dos preguntas, es decir, la correlación corregida por errores de medición $\rho(F_1, F_2)$. Saris y Gallhofer (2014: 310) proporcionan una fórmula para corregir la correlación entre las variables observadas $\rho(Y_1, Y_2)$ y obtener la correlación entre las variables latentes $\rho(F_1, F_2)$:

$$\rho(F_1, F_2) = [\rho(Y_1, Y_2) - CMV] / q_1 q_2 \quad (3)$$

donde *CMV* significa Varianza del Método Común (*Common Method Variance*) y se calcula como el producto de los coeficientes de fiabilidad (r_i) y los del efecto de método (m_i) de ambas variables observadas:

$$CMV = r_1 m_1 m_2 r_2 \quad (4)$$

Los coeficientes del efecto de método se pueden calcular como:

$$m_i = \sqrt{(1 - v_i^2)} \quad (5)$$

La ecuación 3 establece que la correlación entre las variables latentes se puede obtener restando el *CMV* a la correlación entre las variables observadas $\rho(Y_1, Y_2)$ y luego dividir por el producto de los coeficientes de calidad de medición de las dos preguntas ($q_1 q_2$).

Se espera *CMV* cuando las dos variables observadas se miden con la misma escala, lo que lleva a una reacción sistemática de los encuestados a la escala. Por ejemplo, en una escala sin un punto medio neutral, algunos encuestados con una verdadera posición neutral pueden seleccionar sistemáticamente la opción más cercana en el lado positivo, mientras que otros seleccionan sistemáticamente la opción más cercana en el lado negativo, y otros se saltan sistemáticamente la pregunta. Por lo tanto, los investigadores pueden esperar una correlación adicional entre las variables observadas, no vinculada al contenido de las preguntas en sí, sino a la reacción sistemática de los encuestados a un método compartido.

Los coeficientes de calidad de medición de las dos preguntas deben estimarse en un paso anterior, por ejemplo, utilizando experimentos MTMM o el *software Survey Quality Predictor* (SQP) 2.1 (Saris et al., 2011), que genera semiautomáticamente predicciones de calidad de medición de preguntas de encuesta utilizando un rico conjunto de datos de experimentos MTMM previos y algoritmos de bosques aleatorios (*random forests*).

Al comparar las correlaciones sin y con corrección por errores de medición en un conjunto de países diferentes (incluido España), mostraremos cómo cambian las con-

² Para ejemplos más complejos, nos referimos a DeCastellarnau y Saris (2014) y Saris y Revilla (2016).

clusiones sustantivas cuando los errores de medición se tienen en cuenta o no.

DATOS

Ronda 8 de la Encuesta Social Europea

La ESS (<http://www.europeansocialsurvey.org/about/faq.html>) es una encuesta internacional realizada en Europa cada dos años desde 2001. La ESS realiza entrevistas cara a cara de aproximadamente una hora y selecciona nuevas muestras transversales para cada ronda. El cuestionario combina una sección central que se repite en cada ronda y módulos rotativos específicos de cada ronda.

El trabajo de campo de la octava ronda se llevó a cabo entre marzo de 2016 y diciembre de 2017 (datos de la octava ronda de la Encuesta Social Europea, 2016). Los tamaños de las muestras oscilan entre 880 (Islandia) y 2.852 (Alemania), estando España en el medio ($N = 1.958$, véase el apén-

dice C). Las tasas de respuesta para la ronda 8 también varían entre países, oscilando entre el 30,6% (Alemania) y el 74,4% (Israel), con una tasa de respuesta del 67,7% en España y una tasa de respuesta media del 55,4% (ESS, 2017).

El experimento MTMM se llevó a cabo en 23 de los países participantes. En países multilingües, la ESS realiza las encuestas en diferentes idiomas (por ejemplo, catalán y español en España). Dado que el idioma puede afectar la calidad de la medición (Saris y Gallhofer, 2014; Zavala-Rojas, 2016), analizamos cada idioma por separado. Sin embargo, el modelo MTMM no se puede estimar para idiomas con un número reducido de observaciones (por ejemplo, catalán en España; consulte el apéndice C para obtener una lista completa). Así pues, analizamos 27 grupos correspondientes a los grupos lingüísticos con tamaño muestral suficiente existentes en cada uno de los países disponibles (grupos país-idioma).

TABLA 1. Preguntas de encuesta incluidas en el experimento MTMM de la ronda 8 de la ESS

Rasgo	Formulación general de las preguntas
Nivel educativo	¿Qué importancia debería darse a tener un buen nivel educativo en la decisión de permitir o no a una persona que ha nacido y vivido siempre fuera de [país], venir a vivir aquí?
Tradición cristiana	¿Qué importancia debería darse a ser de un país de tradición cristiana en la decisión de permitir o no a una persona venir a vivir aquí?
Cualificación laboral	¿Qué importancia debería darse a tener una cualificación de las que [país] necesita en la decisión de permitir o no a una persona venir a vivir aquí?

Fuente: Elaboración propia.

El experimento MTMM

El experimento evalúa tres rasgos medidos cada uno con tres métodos. Los rasgos pretenden medir tres aspectos del concepto complejo «calificación para la entrada

o exclusión de inmigrantes», respectivamente la importancia de tener: 1) un buen nivel educativo, 2) una tradición cristiana³,

³ Para Israel, «cristiana» se sustituye.

y 3) cualificaciones laborales necesarias en el país, para estar cualificado para entrar en dicho país. La tabla 1 presenta el redactado general de cada pregunta⁴.

Con respecto a los métodos, la tabla 2 resume las características que varían entre métodos y proporciona las etiquetas de los puntos finales para cada escala.

TABLA 2. Variación en las características y etiquetas de los puntos finales en cada escala

	M ₁	M ₂	M ₃
Núm. de puntos	11	10	6
Formato	Batería	Preguntas separadas	Batería
Características			
Núm. puntos de referencia fijos	2	2	1
Correspondencia	Alta	Media	Alta
Preguntas en la tarjeta	No	Sí	No
Etiquetas			
Primera categoría	0 Nada importante	1 Nada importante	0 Nada importante
Última categoría	10 Extremadamente importante	10 Extremadamente importante	5 Muy importante

Fuente: Elaboración propia.

Cinco aspectos varían:

- 1) El número de categorías de respuesta: M₁ es una escala impar (11 puntos), mientras que M₂ y M₃ son escalas pares (10 y 6 puntos, respectivamente).
- 2) Preguntas separadas o batería (es decir, varias preguntas que comparten la misma escala se presentan juntas, la escala se repite solo una vez): M₁ y M₃ presentan las preguntas en formato de batería, mientras que M₂ las presenta como preguntas separadas.
- 3) El número de puntos de referencia fijos (es decir, categorías de respuesta que «no establecen ninguna duda sobre la posición del punto de referencia en la escala subjetiva en la mente del encuestado»; Saris y Gallhofer, 2014: 110): M₁

y M₂ presentan dos puntos de referencia fijos, mientras que M₃ presenta solo uno.

- 4) La correspondencia entre los números y las etiquetas verbales en la escala (por ejemplo, 0 representa mejor la idea de «Para nada» que 1): M₁ y M₃ presentan una correspondencia alta mientras que M₂ presenta una correspondencia media.
- 5) La presentación de la pregunta en las tarjetas que se enseñan a los participantes: por lo general, las tarjetas que la ESS proporciona (es decir, las tarjetas que se presentan a los encuestados para brindar ayuda visual a la vez que el entrevistador hace las preguntas) no contienen la pregunta sino solo las opciones de respuesta. En este experimento, en M₂ las preguntas se muestran en las tarjetas, mientras que en M₁ y M₃ no.

⁴ Para el redactado específico de cada método, véase el apéndice D.

Ilustración de las implicaciones para la investigación sustantiva transnacional de no corregir por los errores de medición

Para ilustrar la implicación de no corregir por los errores de medición en la investigación transnacional, comparamos las correlaciones, antes y después de corregirlas por los errores de medición, entre la importancia que se le da a que un individuo venga de tradición cristiana y la importancia que se le da a sus cualificaciones laborales al momento de decidir si alguien nacido, criado y que vive en el exterior debe poder ir a vivir a un país determinado. En aras de la simplicidad, en esta ilustración nos enfocamos únicamente en uno de los métodos. Elegimos M_3 (escala de 6 puntos en formato de batería, con un punto de referencia fijo, alta correspondencia en la escala y que no proporciona la pregunta en la tarjeta que se enseña al participante), porque, para España, presenta una de las calidades más bajas. Ilustramos las implicaciones para ocho países: Alemania, España, Francia, Finlandia, Italia, Noruega, Portugal y Suecia.

RESULTADOS

Nuestros análisis estiman la calidad de medición para 27 grupos país-idioma, tres rasgos

y tres métodos. Dado que presentar todas las 243 estimaciones de calidad de medición no es práctico, primero agregamos todos los países y presentamos los resultados para cada rasgo y método en comparación con España. Luego, agregamos los rasgos y presentamos los resultados para cada país y método. De esa manera, podemos comparar la calidad de las mediciones, primero, entre rasgos y, segundo, entre países. Finalmente, presentamos un ejemplo de las implicaciones sustantivas de no corregir los errores de medición al comparar España con otros siete países.

Calidad promedio en todos los grupos de países e idiomas, por rasgo y método

La tabla 3 presenta la calidad de medición en España, así como la calidad media, mínima y máxima en los otros 26 grupos país-idioma (excluido España) para los diferentes rasgos y métodos.

Para todos los grupos país-idioma, rasgos y métodos, la calidad más alta obtenida es 0,99 (tradición cristiana- M_1) mientras que la más baja es 0,39 (cualificaciones laborales- M_1). Esto significa que entre el 1% (tradición cristiana- M_1) y el 61% (cualificaciones laborales- M_1) de la varianza en las respuestas observadas proviene de errores de medición.

TABLA 3. Calidad de medición (q^2) en España y media, mínima y máxima calidad de medición de los otros 26 grupos país-idioma, por rasgo y método

Calidad q^2	Nivel educativo			Tradición cristiana			Cualificación laboral			Media de los rasgos		
	M_1	M_2	M_3	M_1	M_2	M_3	M_1	M_2	M_3	M_1	M_2	M_3
Media 26 grupos	0,73	0,64	0,72	0,83	0,71	0,75	0,76	0,68	0,72	0,77	0,68	0,73
Máximo 26 grupos	0,90	0,85	0,87	0,99	0,85	0,92	0,92	0,86	0,85	0,90	0,82	0,85
Mínimo 26 grupos	0,56	0,41	0,41	0,41	0,57	0,64	0,39	0,54	0,42	0,53	0,54	0,49
España	0,85	0,69	0,64	0,87	0,73	0,70	0,88	0,72	0,64	0,87	0,71	0,66

Nota: Las estimaciones de calidad toman valores entre 0 y 1, representando 1 una relación perfecta entre la respuesta observada y el concepto de interés latente.

Fuente: Elaboración propia.

Asimismo, M_1 presenta una calidad más alta en promedio para los 26 grupos país-idioma y para España, para todos los rasgos. Sin embargo, existen algunas diferencias entre España y la media de los otros 26 grupos. Primero, aunque M_1 es el que tiene un mejor rendimiento en ambos casos, las estimaciones de calidad en España son especialmente buenas para todos los rasgos. Las estimaciones de calidad también son más altas para M_2 en España que la media de todos los demás grupos país-idioma. Sin embargo, para M_3 , la calidad en España está por debajo de la media de todos los demás grupos país-idioma. Por otro lado, en España, M_2 presenta una mayor calidad que M_3 en todos los rasgos, mientras que para la media de los demás grupos país-idioma la tendencia es opuesta. Por lo tanto, aunque los formatos de batería pueden sufrir del fenómeno de no diferenciación (Saris y Gallhofer, 2014), en general recomendamos utilizar una escala de 11 puntos presentada en formato de batería, con dos puntos de referencia fijos, alta correspondencia entre números y etiquetas verbales, y ninguna pregunta en la tarjeta en lugar de los otros dos métodos, para medir los tres indicadores estudiados para el concepto «calificación de ingreso o exclusión de inmigrantes».

En cuanto a las diferencias entre rasgos, la «tradición cristiana» alcanza la calidad de medición promedio más alta para todos los métodos, para España y en promedio para los demás grupos de países e idiomas. Esto es interesante ya que se podría pensar que sería el rasgo con mayor propensión a generar sesgos de deseabilidad social, considerándose la religión un tema delicado. Finalmente, las diferencias entre rasgos son consistentes para España y la media de los otros grupos de países e idiomas. Aunque España presenta diferentes estimaciones de calidad, la relación entre las estimaciones de calidad y los rasgos es similar.

Calidad promedio en todos los rasgos, por grupo país-idioma y método

A continuación, las diferencias entre países se analizan con más detalle, esta vez agregando a nivel de rasgos. La tabla 4 presenta la calidad promedio en todos los rasgos, por grupo país-idioma y método, así como la posición de cada grupo país-idioma en el *ranking*, para cada método.

Primeramente, la calidad de la medición entre países y métodos varía de 0,53 (Estonia-Ruso- M_1) a 0,88 (Islandia- M_1). Por lo tanto, en todos los métodos y países, la varianza explicada por los errores de medición va del 12% (Islandia- M_1) al 47% (Estonia-Ruso- M_1). Entonces, existen grandes diferencias en la calidad de medición entre los diferentes grupos país-idioma. La tendencia general es que M_1 (escala de 11 puntos en formato de batería, con dos puntos de referencia fijos, alta correspondencia entre números y etiquetas verbales y sin preguntas en las tarjetas) se desempeña mejor que M_2 y M_3 . Además, los países del centro y norte de Europa presentan, en general, una calidad de medición más alta que sus homólogos del este y del sur.

Comparando España con los demás, España tiene la cuarta calidad más alta para M_1 y la décima más alta para M_2 . Sin embargo, para M_3 , España presenta la cuarta calidad más baja. Por lo tanto, existen diferencias importantes entre los métodos para España, los cuales deben tenerse en cuenta. Primero, usar una escala de 11 puntos presentada en un formato de batería, con dos puntos de referencia fijos, alta correspondencia entre números y etiquetas verbales, y ninguna pregunta en la tarjeta funciona mucho mejor en España que en la mayoría de los grupos de países e idiomas. En segundo lugar, una escala de 6 puntos en formato de batería, con un solo punto de referencia fijo, alta correspondencia entre números y etiquetas verbales, y sin preguntas en la tarjeta, se comporta peor en

España que en la mayoría de los grupos de países e idiomas analizados. Esto sugiere que las diferencias culturales y lingüísticas

entre países afectan el tamaño de los errores de medición asociados con diferentes métodos.

TABLA 4. Calidad media (q^2) de los rasgos agregados, por grupo país-idioma y método

	Calidad media de los tres rasgos			Ranking		
	M ₁	M ₂	M ₃	M ₁	M ₂	M ₃
Austria	0,74	0,67	0,79	20	14	7
Bélgica-Neerlandés	0,83	0,70	0,74	9	11	13
Bélgica-Francés	0,87	0,58	0,68	5	25	20
República Checa	0,79	0,69	0,69	13	13	18
Estonia-Estonio	0,77	0,58	0,72	16	24	14
Estonia-Ruso	0,53	0,63	0,69	27	19	19
Finlandia	0,90	0,75	0,74	1	5	12
Francia	0,82	0,54	0,67	11	27	21
Alemania	0,77	0,80	0,76	15	3	10
Gran Bretaña	0,72	0,73	0,71	21	8	15
Hungría	0,70	0,61	0,64	22	22	25
Islandia	0,88	0,82	0,85	2	1	1
Irlanda	0,83	0,61	0,49	10	23	27
Israel-Árabe	0,88	0,80	0,78	3	2	8
Israel-Hebreo	0,76	0,63	0,62	18	20	26
Italia	0,68	0,57	0,84	24	26	2
Lituania	0,69	0,63	0,81	23	18	5
Países Bajos	0,80	0,65	0,77	12	17	9
Noruega	0,84	0,74	0,82	7	6	4
Polonia	0,75	0,73	0,71	19	9	22
Portugal	0,67	0,61	0,75	25	21	11
Rusia	0,66	0,66	0,83	26	15	3
Eslovenia	0,79	0,66	0,66	14	16	24
España	0,87	0,71	0,66	4	10	23
Suecia	0,83	0,74	0,79	8	7	6
Suiza-Francés	0,85	0,69	0,71	6	12	16
Suiza-Alemán	0,76	0,78	0,70	17	4	17

Fuente: Elaboración propia.

Implicaciones sustantivas para la investigación transnacional: una ilustración

Los resultados demuestran que existen diferencias no despreciables entre España y otros grupos país-idioma. Esto puede tener importantes implicaciones en la investigación transnacional cuando no se tienen en cuenta los errores de medición.

La tabla 5 presenta las correlaciones sin y con corrección por errores de medición para cada país, ordenadas de mayor a menor correlación corregida. Además, presenta el *ranking* de cada país (1 significa la correlación más alta) para correlaciones sin y con corrección por errores de medición.

Podemos ver un aumento en las correlaciones al corregir por errores de medición,

excepto en Italia. Para España la correlación pasa de 0,41 a 0,46. Sin embargo, el cambio no es homogéneo entre países: mientras que para Finlandia la correlación aumenta 0,09 puntos, para Italia se reduce en 0,03 puntos. En consecuencia, comparar países que utilizan correlaciones sin corrección en lugar de correlaciones con corrección por errores de medición lleva a conclusiones sustantivas diferentes. En particular, en términos del *ranking*, sin corrección por errores de medición, Italia presenta la cuarta correlación más alta mientras que con corrección, presenta la más baja. Por lo tanto, si los investigadores quisieran comparar la correlación entre «tradición cristiana» y «cualificación laboral» para España e Italia, no corregir por los errores de medición llevaría a conclusiones erróneas.

TABLA 5. Coeficientes de correlación y ranking sin y con corrección

	Correlación		Ranking	
	Sin corrección	Con corrección	Sin corrección	Con corrección
Finlandia	0,47	0,56	1	1
Suecia	0,41	0,47	2	2
España	0,41	0,46	3	3
Noruega	0,39	0,44	5	4
Portugal	0,38	0,43	6	5
Alemania	0,36	0,42	7	6
Francia	0,35	0,40	8	7
Italia	0,40	0,37	4	8

Fuente: Elaboración propia.

DISCUSIÓN Y CONCLUSIONES

Resultados principales

Nuestro principal objetivo ha sido comparar la calidad de medición de las preguntas de las encuestas en España con otros países europeos, ya que la investigación existente centrada en España desde una perspectiva comparativa, aunque relevante, es aún escasa.

En general, para los tres rasgos considerados, encontramos que la calidad de medición varía mucho entre países, desde un promedio (de todos los métodos y rasgos) de 0,62 en Estonia-Ruso a 0,85 en Islandia, lo que significa que en promedio entre el 62% y el 85% de la varianza en las respuestas observadas se debe a los conceptos de interés latentes mientras que del 15% al 38% se debe a errores

de medición. Los países del centro y norte de Europa presentan, en general, una mayor calidad de medición. Esto podría estar relacionado con el hecho de que los países con bajos niveles de colectivismo y corrupción son menos propensos a la deseabilidad social (Rammstedt, Danner y Bosnjak, 2017) y/o a las diferencias lingüísticas.

Además, la calidad de medición de los tres rasgos considerados fue superior para España que para la media de los otros 26 grupos país-idioma analizados. Estos resultados van en línea con investigaciones previas sobre la calidad de medición (Saris *et al.*, 2010; Revilla, Saris y Krosnick, 2014) basadas también en datos de ESS.

Sin embargo, existen diferencias entre los métodos. M_1 presenta un rendimiento especialmente bueno en España en comparación con la mayoría de los demás países, y ocupa el cuarto lugar de los 27 grupos país-idioma. Sin embargo, para M_3 , España presenta la cuarta estimación de calidad más baja. Por tanto, aunque en general España presenta una calidad de medición superior, los investigadores no deberían asumir una calidad superior a la media en España para cualquier método. Al contrario, deberían considerar que algunos métodos pueden funcionar mejor y otros peor en España que en otros países.

No considerar las potenciales diferencias en el tamaño de los errores de medición al comparar España con otros países afecta a las conclusiones de fondo. En primer lugar, en nuestra ilustración, las correlaciones observadas fueron mayormente subestimadas. Además, los *rankings* de países con mayor y menor correlación con y sin corrección difirieron substancialmente. En particular, sin corrección, España e Italia presentaron correlaciones similares. Esto sugiere que, para ambos países, la creencia de que venir de tradición cristiana es importante para que los inmigrantes estén cualificados para ingresar al país esta correlacio-

nada en similar medida con la creencia de que los inmigrantes que presentan cualificaciones laborales necesarias en dicho país están más cualificados para su ingreso. Sin embargo, la correlación corregida por errores de medición fue mayor para España que para Italia, lo que apunta a una conclusión sustantiva diferente: aunque España e Italia tienen un nivel similar de religiosidad (Evans y Baronavski, 2018), la relación entre la proveniencia de una tradición cristiana y las cualificaciones laborales es más fuerte para España. Esta diferencia entre las correlaciones con y sin corrección entre España e Italia se relaciona principalmente con la diferencia en la calidad de medición de ambos países (0,18 puntos menos en España que en Italia), y en menor medida a la diferencia en CMV (0,04 más alto en España que en Italia). Después de corregir utilizando la ecuación 3, España presenta una correlación verdadera más alta que Italia. Esto indica que, aunque España presenta un CMV algo más alto, la calidad notablemente más baja en España llevó a una subestimación de la correlación en comparación con la sobreestimación de Italia.

Límites y futura investigación

Estos resultados presentan algunas limitaciones. Primero, estos hallazgos son específicos para los temas analizados y los métodos utilizados y no deben ser generalizados a otras preguntas o métodos. En segundo lugar, debido al reducido tamaño de la muestra, no se pudieron analizar algunos idiomas. En particular, no hemos podido utilizar a los encuestados de habla catalana, lo que no permite comparar las estimaciones de calidad entre los idiomas de España. Sin embargo, considerando que otros países presentan diferentes calidades de medición en función del idioma de administración, podríamos esperar lo mismo para España. Una investigación adicional podría explorar específicamente las diferencias entre el catalán y el

español. Además, los experimentos MTMM no son los más adecuados para explicar por qué en algunos países la calidad es más alta que en otros. Futuras investigaciones podrían centrarse en encontrar explicaciones. También, solo hemos ilustrado cómo corregir las correlaciones entre dos conceptos simples para errores de medición. Sin embargo, la corrección por errores de medición se puede aplicar a modelos más complejos (por ejemplo, regresiones). Nos referimos a DeCastellarnau y Saris (2014), Saris y Gallhofer (2014) y Saris y Revilla (2016) para obtener ejemplos y pautas sobre cómo hacerlo para otros modelos. Asimismo, la calidad de la medición proporciona información sobre relaciones estandarizadas. Los investigadores interesados en comparar relaciones no estandarizadas deben estudiar la equivalencia de medición de constructos entre países (Davidov *et al.*, 2014). Finalmente, la estimación del tamaño de los errores de medición también puede verse afectada por errores. Así, incluso las correlaciones corregidas presentan algunos errores.

Para poder sacar conclusiones generales, futuras investigaciones deben explorar nuevos temas y métodos para ver si la tendencia es la misma para diferentes rasgos y escalas. Sin embargo, no siempre es posible realizar experimentos MTMM. Una alternativa es utilizar el software SQP. Utilizando predicciones de SQP, los investigadores podrían obtener una imagen más clara del efecto de diferentes métodos para diferentes preguntas (DeCastellarnau y Revilla, 2017). Asimismo, la sensibilidad de los análisis se podría testear utilizando el software SQP para explorar si las predicciones y estimaciones son similares y, de no ser así, cómo las diferencias afectan las correcciones por errores de medición.

Recomendaciones prácticas

Primero, basándonos en nuestros resultados, para medir el concepto «calificación para la entrada o exclusión de inmigran-

tes» en España recomendamos utilizar M_1 (escala de 11 puntos en formato de batería, con dos puntos de referencia fijos, alta correspondencia entre números y etiquetas verbales y ninguna pregunta en las tarjetas que se enseñan a los participantes) en lugar de M_2 y M_3 .

Segundo, este estudio de caso ilustra lo que previamente se había afirmado en otras investigaciones (p. ej., Saris y Gallhofer, 2007), es decir, que: 1) los investigadores sustantivos deben tener en cuenta que la comparación de relaciones estadísticas estandarizadas entre países solo es posible si la calidad de la medición es la misma, y 2) incluso en este caso, es necesario corregir por los errores de medición para estimar adecuadamente las relaciones de interés; es decir, las que existen entre los conceptos, y no entre las variables observadas, que son solo medidas imperfectas de los conceptos de interés. Por lo tanto, de acuerdo con investigaciones anteriores y los resultados de este nuevo estudio, recomendamos corregir por los errores de medición siempre que sea posible.

BIBLIOGRAFÍA

- Alwin, Duane F. (2007). *Margins of Error: A Study of Reliability in Survey Measurement*. Hoboken, New Jersey: John Wiley and Sons, Inc.
- Andrews, Frank M. (1984). «Construct Validity and Error Components of Survey Measures: A Structural Modelling Approach». *Public Opinion Quarterly*, 48(2): 409-442. doi: 10.1086/268840
- Beilmann, Mai; Kööts-Ausmees, Liisi y Realo, Anu (2018). «The Relationship Between Social Capital and Individualism-Collectivism in Europe». *Social Indicators Research*, 137: 641-664. doi: 10.1007/s11205-017-1614-4
- Beullens, Koen; Loosveldt, Geert; Vandenplas, Caroline y Stoop, Ineke (2018). «Response Rates in the European Social Survey: Increasing, Decreasing or a Matter of Fieldwork Efforts?». *Survey Methods: Insights from the Field*. doi: 10.13094/SMIF-2018-00003

- Bosch, Oriol J.; Revilla, Melanie; DeCastellarnau, Anna y Weber, Wiebke (2019). «Measurement Reliability, Validity and Quality of Slider versus Radio Button Scales in an Online Probability-Based Panel in Norway». *Social Science Computer Review*, 37(1): 119-132. doi: 10.1177/0894439317750089
- Browne, Michael W. (1984). «The Decomposition of Multitrait-Multimethod Matrices». *British Journal of Mathematical and Statistical Psychology*, 37(1): 1-21. doi: 10.1111/j.2044-8317.1984.tb00785.x
- Campbell, Donald T. y Fiske, Donald W. (1959). «Convergent and Discriminant Validation by the Multitrait-Multimethod Matrices». *Psychological Bulletin*, 56(2): 81-105. doi: 10.1037/h0046016
- Corten, Irmgard W.; Saris, Willem E.; Coenders, Germà; Veld, William M. van der; Aalberts, Chris E. y Cornelis, Charles (2002). «Fit of Different Models for Multitrait-Multimethod Experiments». *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2): 213-232. doi: 10.1207/S15328007SEM0902_4
- Couper, Mick P. y Leeuw, Edith D. de (2003). «Non-response in Cross-Cultural and Cross-National Surveys». En: Harkness, J.; Vijver, F. van de y Mohler, P. (eds.). *Cross-Cultural Survey Methods*. New York: Wiley.
- Cudeck, Robert (1989). «Analysis of Correlation Matrices Using Covariance Structure Models». *Psychological Bulletin*, 105(2): 317-327. doi: 10.1037/0033-2909.105.2.317
- Davidov, Eldad; Meuleman, Bart; Cieciuch, Jan; Schmidt, Peter y Billiet, Jaak (2014). «Measurement Equivalence in Cross-National Research». *Annual Review of Sociology*, 40(1): 55-75. doi: 10.1146/annurev-soc-071913-043137
- DeCastellarnau, Anna y Saris, Willem E. (2014). «A Simple Way to Correct for Measurement Errors in Survey Research». *European Social Survey Education Net (ESS Edunet)*. *European Social Survey Education Net (ESS EduNet)*. Disponible en: <http://essedunet.nsd.uib.no/cms/topics/measurement/>
- DeCastellarnau, Anna y Revilla, Melanie (2017). «Two approaches to evaluate measurement quality in online surveys: An application using the norwegian citizen panel». *Survey Research Methods*, 11(4): 415-433. doi: 10.18148/srm/2017.v11i4.7226
- DeMaio, Theresea J. (1984). «Social Desirability in Survey Measurement: A Review». En: Turner, C. F. y Martin, E. (eds.). *Surveying Subjective Phenomena*. New York: Russell Sage.
- ESS (2016). *Data File Edition 2.1. NSD —Norwegian Centre for Research Data, Norway— Data Archive y Distributor of ESS Data for ESSERIC*. Disponible en: <https://www.europeansocialsurvey.org/data/download.html?r=8>, acceso el 21 de septiembre de 2020.
- ESS (2017). *ESS8 - 2016 Fieldwork Summary and Deviations*. Disponible en: https://www.europeansocialsurvey.org/data/deviations_8.html, acceso el 19 de septiembre de 2020.
- ESS (2019). *ESS User Statistics*. Disponible en: http://www.europeansocialsurvey.org/docs/data_users/ESS_data_user_stats_aug_2019.pdf, acceso el 21 de septiembre de 2020.
- Evans, Jonathan y Baronavski, Chris (2018). *How Do European Countries Differ in Religious Commitment?* Disponible en: <https://www.pewresearch.org/fact-tank/2018/12/05/how-do-european-countries-differ-in-religious-commitment/>, acceso el 21 de septiembre de 2020.
- GESIS (2019). *Overview of Comparative Surveys Worldwide*. Disponible en: www.gesis.org/ComparativeSurveyOverview, acceso el 21 de septiembre de 2020.
- Herk, Hester van; Poortinga, Ype H. y Verhallen, Theo M. M. (2004). «Response Styles in Rating Scales: Evidence of Method Bias in Data from Six EU Countries». *Journal of Cross-Cultural Psychology*, 35(3): 346-360. doi: 10.1177/0022022104264126
- Hox, Joop J. y Bechger, Timo M. (1998). «An Introduction to Structural Equation Modeling». *Family Science Review*, 11: 354-373.
- Johnson, Timothy P. y Vijver, Fons J. R. van de (2003). «Social Desirability in Cross-Cultural Research». En: Vijver, F. van de; Mohler, P. y Wiley, J. (eds.). *Cross-Cultural Survey Methods*. Hoboken, New Jersey: Wiley-Interscience.
- Jöreskog, Karl G. y Sörbom, Dag (1996). *LISREL 8: User's Reference Guide*. Uppsala: Scientific Software International.
- Leung, Kwok; Au, Yuk-Fai; Fernández Dols, José Miguel e Iwawaki, Saburo (1992). «Preference for Methods of Conflict Processing in Two Collectivist Cultures». *International Journal of Psychology*, 27(2): 195-209. doi: 10.1080/00207599208246875
- Liao, Pei-Shan; Saris, Willem E. y Zavala-Rojas, Diana (2019). «Cross-National Comparison of Equivalence and Measurement Quality of Response Scales in Denmark y Taiwan». *Journal of Official Statistics*, 35(1): 117-135. doi: 10.2478/jos-2019-0006

- Meurs, A. van y Saris, Willem E. (1990). «Memory Effects in MTMM Studies». En: Saris, W. E. y Meurs, A. van (eds.). *Evaluation of Measurement Instruments by Meta-Analysis of Multitrait-multimethod Studies*. Amsterdam: North-Holland.
- Rammstedt, Beatrice; Danner, Daniel y Bosnjak, Michael (2017). «Acquiescence Response Styles: A Multilevel Model Explaining Individual-Level y Country-Level Differences». *Personality and Individual Differences*, 107(1): 190-194. doi: 10.1016/j.paid.2016.11.038
- Revilla, Melanie y Saris, Willem E. (2013). «The Split-Ballot Multitrait-Multimethod Approach: Implementation and Problems». *Structural Equation Modeling: A Multidisciplinary Journal*, 20(1): 27-46. doi: 10.1080/10705511.2013.742379
- Revilla, Melanie; Saris, Willem E. y Krosnick, Jon A. (2014). «Choosing the Number of Categories in Agree-Disagree Scales». *Sociological Methods & Research*, 43(1): 73-97. doi: 10.1177/0049124113509605
- Revilla, Melanie; Bosch, Oriol J. y Weber, Wiebke (2019). «Unbalanced 3-Group Split-Ballot Multitrait-Multimethod Design?». *Structural Equation Modeling*, 26(3): 437-447. doi: 10.1080/10705511.2018.1536860
- Saris, Willem E. y Andrews, Frank M. (1991). «Evaluation of Measurement Instruments Using a Structural Modeling Approach». En: Biemer, P.; Groves, R.; Lyberg, L.; Mathiowetz, N. y Sudman, S. (eds.). *Measurement Errors in Surveys*. New York: John Wiley and Sons, Inc.
- Saris, Willem E. y Aalberts, Chris (2003). «Different Explanations for Correlated Disturbance Terms in MTMM Studies». *Structural Equation Modeling: A Multidisciplinary Journal*, 10(2): 193-213. doi: 10.1207/S15328007SEM1002_2
- Saris, Willem E. y Gallhofer, Irmtraud N. (2014 [2007]). *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. Hoboken, New Jersey: John Wiley and Sons, Inc.
- Saris, Willem E. y Revilla, Melanie (2016). «Correction for Measurement Errors in Survey Research: Necessary and Possible». *Social Indicators Research*, 127(3): 1005-1020. doi: 10.1007/s11205-015-1002-x
- Saris, Willem E.; Satorra, Albert y Coenders, Germa (2004). «A New Approach to Evaluating the Quality of Measurement Instruments: The Split-Ballot MTMM Design». *Sociological Methodology*, 34(1): 311-347. doi: 10.1111/j.0081-1750.2004.00155.x
- Saris, Willem E.; Satorra, Albert y Veld, William M. van der (2009). «Testing Structural Equation Models or Detection of Misspecifications?». *Structural Equation Modeling: A Multidisciplinary Journal*, 16(4): 561-582. doi: 10.1080/10705510903203433
- Saris, Willem E.; Revilla, Melanie; Krosnick, Jon A. y Eric M. Shaffer (2010). «Comparing Questions with Agree/Disagree Response Options to Questions with Item-Specific Response Options». *Survey Research Methods*, 4(1): 61-79. doi: 10.18148/srm/2010.v4i1.2682
- Saris, Willem E.; Oberski, Daniel L.; Revilla, Melanie; Zavala-Rojas, Diana; Lilleoja, Laur; Gallhofer, Irmtraud, N. y Gruner, Thomas (2011). *The Development of the Program SQP 2.0 for the Prediction of the Quality of Survey Questions*. (RECSM Working Paper). Disponible en: http://www.upf.edu/survey/_pdf/RECSM_wp024.pdf
- Transparency International (2019). *Corruption Perceptions Index 2019*. Disponible en: <https://www.transparency.org/en/cpi/2019>, acceso el 21 de septiembre de 2020.
- Veld, William M. van der; Saris, Willem E. y Satorra, Albert (2008). *Judgement Rule Aid for Structural Equation Models*. (Versión 3.0.4 Beta).
- Zavala-Rojas, Diana (2016). *Measurement Equivalence in Multilingual Comparative Survey Research*. Barcelona: Universitat Pompeu Fabra. [Tesis doctoral].

RECEPCIÓN: 19/11/2019

REVISIÓN: 06/05/2020

APROBACIÓN: 11/09/2020

APÉNDICE A: EJEMPLO DEL CÓDIGO BASE DE LISREL

Analysis Group 1

```
Data ng=3 ni=9 no=221 ma=cm
km file=ATGER-group1.corr
mean file=ATGER-group1.mean
sd file=ATGER-group1.sd
model ny=9 ne=9 nk=6 ly=fu,fi te=sy,fi ps=sy,fi be=fu,fi ga=fu,fi ph=sy,fi
value 1 ly 1 1 ly 2 2 ly 3 3 ly 4 4 ly 5 5 ly 6 6 te 7 7 te 8 8 te 9 9
fr te 1 1 te 2 2 te 3 3 te 4 4 te 5 5 te 6 6
value 0 ly 7 7 ly 8 8 ly 9 9
fr ga 1 1 ga 2 2 ga 3 3 ga 4 1 ga 5 2 ga 6 3 ga 7 1 ga 8 2 ga 9 3
value 1 ga 1 4 ga 2 4 ga 3 4 ga 4 5 ga 5 5 ga 6 5 ga 7 6 ga 8 6 ga 9 6 ph 1 1 ph 2 2 ph 3 3
fr ph 2 1 ph 3 1 ph 3 2 ph 4 4 ph 5 5 ph 6 6
start .5 all
out mi iter= 300 adm=off sc
```

Analysis Group 2

```
Data ni=9 no=219 ma=cm
km file=ATGER-group2.corr
mean file=ATGER-group2.mean
sd file=ATGER-group2.sd
model ny=9 ne=9 nk=6 ly=fu,fi te=sy,fi ps=in be=in ga=in ph=in
fr te 4 4 te 5 5 te 6 6 te 7 7 te 8 8 te 9 9
va 1 ly 4 4 ly 5 5 ly 6 6 ly 7 7 ly 8 8 ly 9 9 te 1 1 te 2 2 te 3 3
equal te 1 4 4 te 4 4
equal te 1 5 5 te 5 5
equal te 1 6 6 te 6 6
value 0 ly 1 1 ly 2 2 ly 3 3
out mi iter= 300 adm=off sc
```

Analysis Group 3

```
Data ni=9 no=228 ma=cm
km file=ATGER-group3.corr
mean file=ATGER-group3.mean
sd file=ATGER-group3.sd
model ny=9 ne=9 nk=6 ly=fu,fi te=sy,fi ps=in be=in ga=in ph=in
fr te 1 1 te 2 2 te 3 3 te 7 7 te 8 8 te 9 9
va 1 ly 1 1 ly 2 2 ly 3 3 ly 7 7 ly 8 8 ly 9 9 te 4 4 te 5 5 te 6 6
equal te 1 1 1 te 1 1
equal te 1 2 2 te 2 2
equal te 1 3 3 te 3 3
equal te 2 7 7 te 7 7
equal te 2 8 8 te 8 8
equal te 2 9 9 te 9 9
value 0 ly 4 4 ly 5 5 ly 6 6
out mi iter= 300 adm=off sc
```

APÉNDICE B: MODIFICACIONES DEL ANÁLISIS DEL MODELO *SPLIT BALLOT-TRUE SCORE-MTMM*, AJUSTE DEL MODELO Y EVALUACIÓN DE JRULE

Grupo país-idioma	Modificaciones del modelo (notación LISREL)	df	χ^2	Núm. de errores de especificación por JRULE
Austria	Free GA 2 4 TE 8 8 TE 5 5	36	78,90	0
Bélgica-Neerlandés	Free TE 1 1	38	92,20	0
Bélgica-Francés	Free GA 8 6	38	60,59	2
República Checa	Free TE 7 7 TE 2 2 TE 5 5 GA 5 5	35	86,50	0
Estonia-Estonio	Free TE 6 6 TE 5 5 TE 8 5	36	95,20	0
Estonia-Ruso	Free TE 9 9	38	55,87	4
Finlandia	Free TE 8 8 TE 7 7 TE 4 4 GA 5 5 TE 8 2	34	91,10	2
Francia	Free TE 4 4 TE 7 7 GA 4 1 TE 3 1	35	77,32	3
Alemania	Free TE 7 7 TE 2 2 TE 4 4 TE 6 6 GA 2 4	34	79,69	3
Gran Bretaña	Free TE 4 4 TE 1 1 TE 3 3 TE 6 6 GA 8 6 GA 2 4 GA 6 5	32	83,00	4
Hungría	Free TE 4 4 TE 7 7 TE 5 5 TE 9 9 GA 7 6 (G2)	34	83,27	1
Irlanda	Free TE 7 7 TE 8 8 TE 4 4 GA 7 6 GA 8 6 GA 9 6 (G2); fix TE 2 2 (va 0)	34	88,41	2
Israel-Árabe	Free TE 8 8 TE 8 8 GA 2 4 PH 5 4 GA 9 3 (G3)	34	41,70	1
Israel-Hebreo	Free TE 1 1 GA 8 6 GA 5 5	36	81,45	2
Islandia	Free TE 4 4 TE 6 6 TE 1 1	36	71,45	0
Italia	Free TE 4 4 TE 7 7 TE 5 5 PH 5 4	37	98,26	1
Lituania	PH 5 4	38	74,01	0
Países Bajos	TE 6 6 TE 4 4 GA 5 5 GA 6 5 PH 5 4	36	58,20	1
Noruega	TE 4 4 TE 8 8 TE 7 7 GA 2 4	37	88,08	2
Polonia	TE 4 4 TE 1 1 TE 7 7 TE 7 1	37	85,50	0
Portugal	Ninguna	39	69,03	0
Rusia	TE 4 4 TE 5 5 TE 2 2 TE 1 1 GA 6 5 (G2) PH 5 4	33	83,35	1
Eslovenia	TE 1 1	38	60,48	3
España	TE 4 4 GA 8 6 PH 6 5	36	92,04	0
Suecia	TE 1 1 TE 7 7 TE 2 2 TE 4 4 GA 5 5	36	68,07	2
Suiza-Francés	TE 4 4 TE 6 6 TE 3 3	36	46,95	2
Suiza-Alemán	TE 4 4 TE 7 7 GA 8 6	36	87,70	2

Fuente: Elaboración propia.

APÉNDICE C: TAMAÑO MUESTRAL DE LA RONDA 8 DE LA ESS POR GRUPO PAÍS-IDIOMA

Grupo país-idioma	Grupo 1	Grupo 2	Grupo 3	Casos totales
Austria	221	219	228	668
Bélgica-Neerlandés	345	365	342	1.052
Bélgica-Francés	236	241	237	714
República Checa	763	777	726	2.266
Estonia-Estonio	522	510	520	1.552
Estonia-Ruso	169	139	159	467
Finlandia	608	603	590	1.801
Francia	685	680	696	2.061
Alemania	959	958	935	2.852
Gran Bretaña	662	653	644	1.959
Hungría	527	557	530	1.614
Irlanda	913	933	911	2.757
Israel-Árabe	175	168	182	525
Israel-Hebreo	660	663	685	2.008
Islandia	295	293	292	880
Italia	935	843	848	2.626
Lituania	654	662	627	1.943
Países Bajos	557	554	570	1.681
Noruega	497	563	485	1.545
Polonia	591	585	516	1.692
Portugal	408	451	411	1.270
Rusia	822	812	796	2.430
Eslovenia	447	424	436	1.307
España	648	599	590	1.837
Suecia	541	506	504	1.551
Suiza-Francés	123	133	129	385
Suiza-Alemán	361	367	350	1.078

Nota: Finlandia-Sueco, Israel-Ruso, Lituania-Ruso, España-Catalán y Suiza-Italiano no se pudieron analizar debido a que el tamaño de la muestra era < 100 casos por grupo.

Fuente: Elaboración propia.

APÉNDICE D: FORMULACIONES DE LAS PREGUNTAS EN EL CUESTIONARIO, POR MÉTODO

Introducción (similar en todos los casos)

Hay personas de otros países que vienen a vivir a [país] por distintas razones. Algunas tienen antepasados que eran de aquí. Otras vienen a trabajar o para reunirse con sus familias. Otras vienen porque están amenazadas. Aquí hay algunas preguntas sobre este tema.

Método 1

¿Qué importancia debería tener cada uno de los siguientes aspectos en la decisión de permitir o no a una persona que ha nacido y vivido siempre fuera de España, venir a vivir aquí? En primer lugar, ¿qué importancia debería tener que esa persona... **Leer cada frase...**

	Nada importante					Extremadamente importante					(NC)	(NS)			
C33	...	tenga un buen nivel educativo?	00	01	02	03	04	05	06	07	08	09	10	77	88
C34	...	sea de un país de tradición cristiana?	00	01	02	03	04	05	06	07	08	09	10	77	88
C35	...	tenga una cualificación laboral de las que [país] necesita?	00	01	02	03	04	05	06	07	08	09	10	77	88

Método 2

¿Qué importancia debería darse a tener un buen nivel educativo en la decisión de permitir o no a una persona que ha nacido y vivido siempre fuera de [país], venir a vivir aquí?

Nada importante					Extremadamente importante					(NC)	(NS)
01	02	03	04	05	06	07	08	09	10	77	88

¿Qué importancia debería darse a ser de un país de tradición cristiana en la decisión de permitir o no a una persona venir a vivir aquí?

Nada importante					Extremadamente importante					(NC)	(NS)
01	02	03	04	05	06	07	08	09	10	77	88

¿Qué importancia debería darse a tener una cualificación de las que [país] necesita en la decisión de permitir o no a una persona venir a vivir aquí?

Nada importante					Extremadamente importante					(NC)	(NS)
01	02	03	04	05	06	07	08	09	10	77	88

Método 3

Tarjeta 30 ¿Qué importancia debería tener cada uno de los siguientes aspectos en la decisión de permitir o no, a una persona que ha nacido y vivido siempre fuera de [país], venir a vivir aquí? En primer lugar, ¿qué importancia debería tener que esa persona... **leer cada frase...**

		Nada importante			Muy importante			(NC)	(NS)
C39	... tenga un buen nivel educativo?	00	01	02	03	04	05	7	8
C40	... sea de un país de tradición cristiana?	00	01	02	03	04	05	7	8
C41	...tenga una cualificación laboral de las que [país] necesita?	00	01	02	03	04	05	7	8