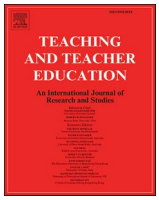
Teaching and Teacher Education 146 (2024) 104631

Contents lists available at ScienceDirect

Teaching and Teacher Education

journal homepage: www.elsevier.com/locate/tate



Research paper

Unveiling teachers' work preferences: A conjoint experiment on the implications of school governance reform across three countries

Antonina Levatino [a,*], Gerard Ferrer-Esteban [b], Antoni Verger [c]

[a] Department of Didactics and Educational Organization, Faculty of Education, Universitat de Barcelona, Pg. Vall d'Hebron, 171, Edifici Llevant, 08035, Barcelona, Spain
[b] Faculty of Psychology and Education Sciences, Universitat Oberta de Catalunya, Rambla del Poblenou, 156, 08018, Barcelona, Spain
[c] Department of Sociology, Universitat Autònoma de Barcelona, Edifici B - Campus Bellaterra, 08193, Cerdanyola, Barcelona, Spain

ARTICLE INFO

Keywords:
Teachers' work
Teachers' preferences
School policy
School governance reforms
Accountability
Conjoint experiment

ABSTRACT

School governance reforms have changed teachers' work in many aspects and have been associated with increasing teachers' discontent and demotivation. Research on which school policies and organizational practices teachers prefer is scarce and faces challenges. Our conjoint experiment identifies the importance given by teachers to different work dimensions altered by recent reforms and teachers' preferences regarding school policies in three contexts. Internationally-shared preferences include qualitative teaching assessments, socially mixed classes, clear goal-setting, and collective rewards. Context-specific preferences include individual incentives in Chile, a low-pay, high-stakes accountability system, and peer support in Norway and Catalonia, which are systems with collaborative governance traditions.

1. Introduction

Ensuring that teachers are motivated and fully engaged in their work is crucial for achieving educational success. Teachers' work motivation encompasses the "energetic forces" that not only attract individuals to the teaching profession but also influence the time, dedication and effort they devote to their roles (Kanfer et al., 2017; Pinder, 2008: 11). Motivation and professional engagement are closely interconnected phenomena that significantly impact the sense of fulfillment and satisfaction teachers derive from their work (Ainley & Carstens, 2018). However, many educational systems are currently struggling with a rising tide of teachers' discontent, demoralization, and dissatisfaction, leading to significant challenges such as teacher shortages, attrition, and burnout (Admiral & Kittelsen Røberg, 2023; Harmsen et al., 2018; Kraft & Lyon, 2022; Räsänen et al., 2020). These declines in teachers' morale, extending beyond mere workforce issues, have far-reaching implications for educational attainment and the quality of educational systems, given the established relationship between teachers' motivation, well-being, and performance (De Clercq et al., 2022; Lazarides & Schiefele, 2021; Shoshani & Eldor, 2016). Against this backdrop, devising policy frameworks that can effectively reinvigorate teachers' engagement with their work emerges as a crucial challenge.

Existing research consistently shows that interpersonal relations and conditions in the workplace, such as supervisor support, peer relations, teachers' assessments, and the level of autonomy, significantly affect work motivation and satisfaction (Admiral & Kittelsen Røberg, 2023; Kanfer et al., 2017; Latham & Pinder, 2005). Recent educational reforms have notably influenced these and other teachers' working conditions, though not always positively. Widely adopted school reforms, such as those that have embraced the tenets of New Public Management (NPM), have shaped teachers' work by introducing new organizational and evaluation methods, applying managerial principles in the schools, and encouraging competitive behaviors among schools. Existing literature has shown that NPM-like policy instruments have also transformed the work relationships within schools (Ball & Maroy, 2009) and been associated with growing teachers' unrest, demoralization and pressure, as they are viewed as diminishing their agency and control over core educational processes (Salokangas & Wermke, 2020).

Given this background and the polarized debate surrounding these reforms in education and other public sectors (see Lapuente & Van de Walle, 2020), it is crucial to explore teachers' preferences regarding key work dimensions affected by contemporary school reforms, such as evaluation methods, incentives, goal-setting approaches, and teachers' support. To our knowledge, not much research has however delved into this topic. A pioneering attempt to explore teachers' preferences and work motives in relation to contemporary school reforms has been

* Corresponding author.
E-mail addresses: Antonina.levatino@ub.edu (A. Levatino), gferrere@uoc.edu (G. Ferrer-Esteban), antoni.verger@uab.cat (A. Verger).

https://doi.org/10.1016/j.tate.2024.104631
Received 21 January 2023; Received in revised form 2 May 2024; Accepted 4 May 2024
Available online 17 May 2024

conducted by Mintrop and Ordenes (2017) using self-reported preferences of teachers in the context of the US. However, as the authors admit, self-reporting and direct questioning might be affected by social desirability biases. Our study uses a conjoint experimental approach, a method successfully applied in various disciplines such as social and political sciences and marketing research. This approach allows us to explore individual preferences in a way that overcomes the limitations of direct questioning, a process we explain in detail in the methodology section of the manuscript.

In this paper, we analyze and compare data collected in three different policy contexts: Chile, Norway, and Catalonia (Spain). Primary and secondary school teachers from urban areas were asked to choose between pairs of schools, which we refer to as *working scenarios*. These scenarios differ by five characteristics (that, hereinafter, we call *attributes*) related to main work dimensions influenced by contemporary school governance reforms, namely school composition, assessment of teaching quality, goal-setting, relations, outcome-based financial incentives. Our study has three key objectives: 1) to determine which work dimensions, shaped by recent school governance reforms, are more important to teachers; 2) in relation with each of these dimensions, to identify the specific school policies and organizational practices that teachers prefer; and 3) to explore how teachers' preferences across countries align or differ.

This paper thus contributes to existing research in several ways. First, it explores for the first time how teachers consider and evaluate, on the same scale, different school attributes related to the work dimensions most affected by recent school reforms. Second, its experimental approach overcomes the limitations of direct questioning (Hainmueller et al., 2014), enabling an unbiased estimation of the importance that teachers give to different work dimensions. It also mitigates the potential desirability bias at stake when teachers are asked about financial incentives based on productivity, and to other policy instruments that are controversial and often seen as politically charged. Lastly, from a policy standpoint, the comparative approach enables the identification of both similarities and differences among teachers operating within three different policy environments, particularly contrasting in the context of school and teacher policies.

2. Conceptual framework

2.1. School governance reforms and teachers' work

Recent school reforms have significantly altered teachers' work. Arguably, the reforms that have more profoundly transformed the governance of schools over the past decades are those encompassed within the NPM movement (Hansen & Jacobsen, 2016). Emerging in Anglo-American countries in the 1980s to modernize, de-bureaucratize, and improve the transparency and efficiency of the public sector, NPM-informed reforms have progressively penetrated a growing number of countries characterized by diverse political and institutional traditions (Gunter et al., 2016). These reforms encompass a wide range of policies that promote outcomes-based management and hands-on leadership approaches, specifically granting principals more freedom to manage school resources and teachers' work, which implies more hierarchical *relations* and leadership styles within schools (Fitzgerald, 2009). They also encourage the parents' empowerment as users through various mechanisms, including school choice and access to performance data. Within this reform context, performance-based accountability (hereafter PBA) has expanded too, implying the use of results from large-scale standardized tests as the main benchmarks of school and teaching quality improvement (Lingard et al., 2015). Depending on the level of performance, schools and their professionals may face consequences of a different nature, from outcome-based rewards and sanctions to more symbolic consequences – mostly related to the visibility of test results to wider audiences (Levatino, Parcerisa, & Verger, 2024).

Many of these policies, as is the case of PBA, are based on *goal-setting* motivational theories. These theories suggest that clear and challenging goals, coupled with outcome-focused incentives (rewards and/or sanctions), would positively impact employees' performance (Latham & Locke, 2007). The assumption is that by associating students' learning outcomes with teachers' productivity, *outcome-based financial rewards* will be internalized as proxies of educational quality and, thus, trigger extra effort (Harackiewicz & Sansone, 2000; Saban, 2003). Policies like these conceive school actors as benefit-maximizing individuals who will put more effort in promoting students' learning if the right structure of incentives is in place (Ehren et al., 2015). In fact, in countries that have gone the furthest in deploying the accountability agenda, the *assessment of teaching quality* and/or *teachers' job promotion* are increasingly connected to student achievement data, for instance, through value-added and/or growth models (Amrein-Beardsley, 2014).

Contemporary reforms, by promoting school autonomy, are also influencing between-school relations. Increased autonomy is associated with schools promoting different educational approaches and, consequently, the diversification of educational systems. Nonetheless, these policies can alter *schools' student body composition*, as schools may, actively or passively, target different student profiles, often to attract academically skilled students (Altrichter et al., 2014). In contexts with a wide margin of parental choice, public information on school performance can influence school demand, and thus, school composition (Maroy & Voisin, 2017). Hence, in countries where these educational reforms have advanced the most, the educational system is not only more differentiated, but also more socially segmented. This is significant because teaching and teachers' experiences can greatly vary with the student body's demographics. Although education researchers often assume that teachers prefer working in affluent schools (Ingersoll & May 2012), others believe that public-service-oriented teachers are drawn to disadvantaged areas, where they can make a bigger difference (Mintrop & Ordenes, 2017).

2.2. School governance reforms and teachers' motivation

Education research has consistently shown that NPM reforms in education are usually associated with increased pressure to perform (Perryman et al., 2011). The presence of outcome-based rewards and sanctions can increase pressure on teachers to enhance student performance, potentially impacting their teaching motivation (Cuevas et al., 2018). But do teachers like working in environments that thicker hands-on school management and accountability has brought about? Research findings are mixed, but there is growing consensus that high-performance pressure can lead to different side-effects (Zhao, 2017) and undermine teachers' morale (Collins, 2014). Even when outcomes-based financial rewards are not involved, if teachers accept and internalize test results as valid measures of their work quality (Sullivan et al., 2021), they might experience self-imposed pressure. Consequently, to avoid the shame and guilt of failure, they might use instructional strategies they disagree with, leading to emotional and social costs (Braun & Maguire, 2020). Other factors, such as the erosion of *relationships within schools*, decreased support from families and peers, and increased workloads (Gewirtz et al., 2021), have also been identified as consequences of contemporary school governance reforms that significantly affect teachers' enjoyment of their work (Booher-Jennings, 2005; Ryan & Deci, 2020).

Overall, research is inconclusive on the power of incentives and other accountability tools to increase teacher engagement, and on whether they can crowd out intrinsic motivation (Supovitz, 2009).[1] The 'overjustification effect' (Levy et al., 2017) predicts that individuals

[1] In contrast to extrinsic motivation, intrinsic motivation is "self-determined" (Ryan & Deci, 2017), as it stems from personal desires, enjoyable experiences, self-identification with the profession and pro-social commitment, among other factors (Grant, 2007).

might increasingly perceive their engagement as merely the result of external rewards rather than their own enjoyment, or conversely, they might view rewards as a means of exerting control over their work, potentially leading to job aversion (Frey & Jegen, 2002). Along these lines, research has found that, rather than seeing externally defined goals, standards and rewards as motivators, teachers perceive these elements as distortions of their work's true meaning (Mintrop, 2004). Furthermore, studies indicate that while initial cycles of performance-based assessments can generate enthusiasm and productive pressure within schools at the start, these policies tend to lose their motivational effect as they become routine and persist over time (Penninckx et al., 2016).

A pioneering attempt to study teachers' work motivation in relation to contemporary school reforms, specifically regarding how teachers perceive goals, standards, and rewards as motivators, was conducted by Mintrop and Ordenes (2017). These authors relied on teachers' self-reported information on work engagement to show that intrinsic motives such as work-related challenges, pro-social commitment, ownership[2] and pleasure, were ranked as the most important motives, whereas financial incentives had a very weak or negligible impact on work motivation. However, as the authors admit (p. 16), self-reported motives may reflect "desired ideologies, rather than true motivational patterns". Interestingly, the same study reports that a large majority of surveyed teachers found performance-related financial rewards for both teachers and schools acceptable. It could be that, to some extent, admitting to being motivated by financial reward is not socially acceptable among teachers, as the traditional teacher occupational ideology usually emphasizes the profession's social mission over material individual gains (e.g. Lortie, 1975), and the "vocational" nature of teachers' work (e.g. Anthony & Ord, 2008; Manuel & Hughes, 2006). There are thus reasons to hypothesize that due to social desirability biases, teachers might not readily admit that they find extrinsic financial rewards attractive when explicitly asked. Given these considerations, direct questioning has its limitations, which our study aims to address using an experimental approach.

3. Our study: research questions

This state of the art has shown how school governance reforms, such as those informed by NPM, are likely to influence several work dimensions and shape the school settings where teachers operate, with possible repercussions on their work engagement and motivation. As shown above, work dimensions particularly affected by NPM-like reforms include goal-setting, outcome-based financial incentives, assessment of teaching quality, within-school relationships, and school composition. Because of the widespread changes brought about by NPM policies on schools and the teaching profession, we consider it important to explore: 1) the degree of importance teachers assign to different work dimensions most affected by school governance reforms, and 2) teachers' preferences regarding each of these dimensions in countries with distinct policy settings.

More specifically, our study aims at answering three research questions that are key to deepening our understanding of teachers' preferences regarding their work: Which work dimensions, among the ones most affected by NPM reforms, matter most for teachers? What are the attributes – i.e., policies and practices related to the previous dimensions – that teachers prefer? To what extent do teachers working in different countries agree, or disagree, on the importance given to different school attributes?

[2] Defined as the self-identification of one person with the work's aims (see Mintrop & Ordenes, 2017, p. 14).

4. Methodology

4.1. Data source and case selection

The data for this study come from a teacher questionnaire containing the conjoint experiment (Levatino, 2021), administered to a representative sample of primary and lower secondary schools in urban areas of three different policy contexts: Chile, Norway, Catalonia (Spain).[3]

These three policy contexts were selected because all of them have been exposed to NPM ideas, albeit in different periods of time and with different levels of adoption. In these contexts, different policy frameworks regarding school choice, PBA, and school leadership have been enacted. Teachers' policies also vary significantly, which affects professional autonomy, horizontal relations, and working conditions within schools, as shown in Table 1.

4.2. Design of the conjoint experiment

Addressing individual preferences presents methodological challenges. First, asking subjects directly about their preferences, for instance through traditional survey questions, often leads respondents to state most aspects as important (Horng, 2009), impeding an unbiased identification and comparison of the importance given to different aspects on the same scale. Second, asking about single aspects is an exercise that may not reflect real preferences, which usually emerge after the simultaneous consideration of multiple aspects (Bansak et al., 2018). Third, as mentioned before, self-reporting might reflect desirable professional ideologies rather than true motivational patterns, especially for the teaching profession, which has normally been considered 'vocational' and intrinsically motivated (Lortie, 1975). This means that information coming from direct questioning might be affected by social desirability bias. The conjoint experiment approach allows us to overcome all these challenges.

In this survey experimental approach, respondents are presented with two distinct options and are asked to choose one of them (Hainmueller et al., 2014). Each option has different dimensions, while each dimension has several attributes. By forcing the respondents to make decisions between profiles with different attributes, conjoint experiments allow researchers to estimate, in an unbiased way, the relative importance of each dimension and attribute in determining whether a profile is chosen. This gives insight into how respondents value and prioritize different dimensions and attributes (Hainmueller et al., 2014). Moreover, conjoint experiments also mitigate social desirability bias. Indeed, since the options presented have several different attributes, respondents are less concerned about researchers' intent to link individual choices to specific attributes (Bansak et al., 2018).

Introduced in the Seventies in marketing research (Green & Rao, 1971), conjoint analysis has been increasingly used in other disciplines as a tool to analyze multidimensional preferences (e.g. Gallego & Marx, 2017; Hainmueller et al., 2014; Hainmueller & Hopkins, 2015). Recent educational research has applied similar approaches to investigate teachers' preferences for specific working conditions, such as commute time, class size, salary, school facilities, resources for students, health insurance offered, tenure perspective and students' characteristics (e.g. Horng, 2009; Johnston, 2020; Lovison & Mo, 2022; Viano et al., 2021). Horng (2009) and Viano et al. (2021) applied an adaptive conjoint approach: options' attributes were not randomized but depended on

[3] Given the diversity in educational policies across Spanish regions, this study needed to focus on a specific area. We chose Catalonia not only because it is one of Spain's most developed and urbanized regions but also because of its diverse educational landscape. This diversity includes school ownership, with a significant presence of subsidized private schools, and varied educational approaches and teacher policies, making it a particularly relevant case study, as shown in Table 1.

Table 1
School and teacher policies in the three cases under study.

Case	GDP per capita, PPP ($) [1]	School policies			Teacher Policies		
		PBA[a]	Leadership	School choice	Work regulation	Professional Autonomy	Salaries
Norway	114,930	Thick and medium-stakes. Test results used for school improvement.	Horizontal and collegial. Highly collaborative school culture.	Restricted choice. Marginal % of private subsidized schools (1.6%).	Professional knowledge and autonomy-based regulation. Highly qualified status.	High and institutionalized. Teachers actively participate in school decisions (85.6%).	Income above OECD average. Salary satisfaction above average.
Catalonia (Spain)	46,333 [2]	Thin and low-stakes. School results are not published and hardly used by schools.	Horizontal- but recent reforms encourage the principal to become a staff manager.	Controlled choice. High % of private subsidized schools (24.2%) [2].	Bureaucratic and subject matter expertise-based regulation.	High, but not institutionalized (de *facto* autonomy). Teachers participate in school decisions (75.7%) [2].	Income above OECD average. Salary satisfaction above average [2].
Chile	30,209	Thick and high stakes. Test results are published and used to evaluate teachers' effectiveness.	Hierarchical principals decide on teachers' salaries and promotion.	Free choice. Very high % of private subsidized schools (51%).	Market & standards-based regulation. Flexible work arrangements.	Medium-high, restricted by school management. Teachers' participation in school decisions below OECD average (58.9%).	Income below OECD average. Salary satisfaction below average.

[1] Gross Domestic Product (GDP) at purchasing power parity (PPP) per capita.
[2] Spanish average.
[a] In this column, we delineate not only the varying stakes of accountability (high and low) but also the 'thickness' of the accountability system, which pertains to the diversity and number of accountability tools embedded in the system that impact the daily operations of schools.
Source: Authors, based on data from OECD (2016, 2020); Högberg and Lindgren (2021); Voisin and Dumay (2020); World Bank (2024).

respondents' answers to previous questions. The approach used by Lovison and Mo (2022) and Johnston (2020) was experimental, similar to the one applied in this paper. Nonetheless, our research approach and goals differ from theirs in that our focus is not on understanding why teachers decide to join a specific school over another, which could help understand teacher turnover in contexts where teachers have the choice of their workplace. Although we also ask respondents to choose between two hypothetical working scenarios (schools), our goal is to explore teachers' preferences in relation to work dimensions affected by new NPM reforms, so as to foster reflection on how teacher motivation and satisfaction can be built and sustained through public policy.

Our experiment has a choice-based conjoint design. We asked respondents to choose between two alternative working scenarios (schools) that randomly vary on several attributes referring to five work dimensions, which have been directly or indirectly altered by NPM-inspired policies. The instructions read: "*Below we present you with two pairs of schools that differ in some aspects. Please indicate which school you would prefer to work in, if you could choose one of them. Which of the two schools would you prefer?*". As mentioned above, the dimensions were: school composition/students, assessment of teaching quality, goal setting, relations, and external outcome-based financial incentives. To ensure that each attribute of a given dimension appeared with equal frequency, as Table 2 shows, each dimension is made up of the same number of possible attributes (Huber & Zwerina, 1996). The order in which the dimensions appeared to the respondent was also randomized each time to prevent an order effect (Saris & Gallhofer, 2014), that is, systematic differences in responses caused by the sequence in which dimensions are presented. Table 2 presents each dimension, as well as the attributes that were randomly selected for each dimension.

The conjoint experiment was inserted in the teacher questionnaire of the Reformed project (Levatino, 2021), which was administered to a representative sample of primary and lower secondary schools located in urban areas of three contexts under study. Table 3 illustrates an example of a task displayed to the respondents, including the instruction, a random pair of fully randomized scenarios (schools), and the question they were asked to answer.

Each respondent was presented sequentially with two pairs of working scenarios (schools) with fully randomized attributes and was asked to choose between them. Full randomization implies that, in some cases, the working scenarios displayed were completely different, while in others, they differed only in certain attributes. This also means that no combination of attributes was restricted. As is common in conjoint experiments, we employed an independent design, whereby one respondent only visualized two pairs of working scenarios. Indeed, in conjoint experiments, the number of possible combinations of attributes is huge, and "only a small fraction" might ever be observed (Hainmueller & Hopkins, 2015, p. 535), It is sufficient to present a small number of possible scenarios (in some experiments, less than 0.04 percent has been reported) to estimate all attribute-level main effects on an uncorrelated basis (Green et al., 2001: S59). Straightforward estimates largely rely on successful randomization, ensuring orthogonality and balance, as explained in subsection 4.4.

Table 2
Conjoint experiment: dimensions and attributes.

DIMENSION	ATTRIBUTES
School composition (Students)	- Advantaged (easy-to-teach students) - Mixed ability (diversity of learning paces) - Struggling (hard-to-teach students)
Assessment of teaching quality	- Teachers are assessed based on students' national test results - Teachers are assessed based on classroom observation - Teachers are assessed based on teacher's portfolio
Goal-setting	- Goals are well defined and well communicated - Goals are not always clear and well communicated - No performance goals are set
Relational dimension (supportive engagement)	- The principal is engaged and very supportive - Parents are engaged and very supportive - The other teachers are engaged and very supportive
Outcome-based financial rewards	- Yearly salary bonus for individual teachers according to teaching assessment results - Yearly budgetary rewards for the school according to teaching assessments results at the school level - No salary bonuses or budgetary rewards attached to the teaching assessment

Table 3
Conjoint experiment: example of task displayed to the respondents.

Below we present you with two pairs of schools that differ in some aspects. Please indicate which school you would prefer to work in, if you could choose one of them.
Which of the two schools would you prefer?

	School A	School B
School composition (Students)	Advantaged (easy-to-teach students)	Struggling (hard-to-teach students)
Assessment of teaching quality	Teachers are assessed based on teacher's portfolio	Teachers are assessed based on students' national test results
Goal-setting	Goals are not always clear and well communicated	No performance goals are set
Relational dimension	Parents are engaged and very supportive	The other teachers are engaged and very supportive
Outcome-based financial incentives	Yearly salary bonus for individual teachers according to teaching assessment results	Yearly budgetary rewards for the school according to teaching assessments results at the school level
Which school would you prefer?		
School A o		
School B o		

4.3. Challenges of the methodological approach

One challenge of the conjoint approach is the inherent complexity of the task that the respondents complete, which could lead to unmeaningful responses. This complexity can arise from a high number of dimensions and attributes, potentially imposing a high cognitive load on respondents. To address this challenge, we decided to keep the number of the analyzed work dimensions and attributes low. The presentation of two pairs of profiles on each screen also reduces respondents' cognitive bias and the risk of satisficing (Krosnick, 1999). Besides, we ask respondents to answer the choice task twice. Recent research has shown that researchers can request respondents to repeat the task even up to 30 times without being afraid of data quality deterioration (Bansak et al., 2018).

In terms of external validity, the artificial nature of the scenarios could also represent a challenge if the scenarios do not reflect situations that respondents find in their own countries, leading to confusion or disengagement. To address this challenge, we simplified the tasks by displaying fewer attributes (a maximum of three) for each dimension. Additionally, we made an effort to ensure they are plausible and realistic in all the countries involved. On this point, it should also be pointed out that recent empirical research (Hainmueller et al., 2015) has shown that conjoint experiments tend to perform very well (and better than other types of survey experiments) in terms of external validity.

4.4. Teachers' sampling procedure

In order to sample teachers, we employed a three-stage sampling procedure (Ferrer-Esteban, 2023) to ensure representativeness of both schools and teachers in the areas of interest. This sampling strategy was adapted in each context according to feasibility criteria, mainly considering data availability.

The first step was constructing the schools' sampling frame, following the criterion that schools should be located in urban areas. The decision to focus on urban areas was taken according to one of the goals of the Reformed project, which is to analyze how schools respond to PBA within local education spaces, where market dynamics of school choice and between-school competition are found (Jabbar, 2015). The decision to establish whether an area is urban varied by country. In Norway, a country with large, inhabited areas and sparsely dense urban centers, we combined the indicator of population density with the municipality size, as well as whether the urban settlement areas are classified as 'central' by the Norwegian Statistics Bureau. In Chile, we focused on three provinces, in which the municipalities are mainly urban contexts: Santiago, Valparaíso and Concepción. We then excluded those municipalities with a density of less than 150 inhabitants per square kilometer. Finally, in Catalonia (Spain), we combined the classification of large urban areas of the Ministry of Public Works and Transport with the indicator of population density.

In the second sampling stage, schools were drawn with systematic probability proportional to size (PPS), which means sorting the schools of the sample frame within an explicit stratum, by the implicit stratification variables. The sample size of each explicit stratum was determined with a proportional allocation. The sampling strategy was adapted to each geographical context and the availability of data. In Norway, the explicit stratum was defined by school ownership, and the implicit stratification variables were educational levels, geographical area, and school size (number of teachers). In Chile, the explicit stratum was defined by school ownership (public, subsidized, and private schools), and then the implicit stratification variables were municipality and school size (number of teachers). In Catalonia (Spain), the explicit stratum was defined by school ownership (public, subsidized schools), and the implicit stratification variables were educational levels, province, and municipality.

Finally, in the third sampling stage, we employed a random selection of teachers, both subject and classroom teachers, within schools to ensure an unbiased representation and cover all teaching and evaluation approaches. In Chile and Catalonia, we randomly sampled 20 teachers in each school, according to the following eligibility criteria: teachers working in the classroom for more than 10 h per week, teachers with subjects both evaluated and not evaluated by standardized tests, teachers working at least part-time and not on parental leave. The procedure for selecting teachers was as follows: first, we created two spreadsheets. One spreadsheet included a list of all teachers in the school whose subjects were evaluated through standardized tests, while the other included all teachers whose subjects were not tested. Then, a random number was generated for each teacher in both lists. The lists were then sorted in ascending order based on the generated random numbers, and the top ten teachers of each list were selected to form the sample. On the other hand, a random selection of teachers was not possible in Norway; thus, everyone who met the eligibility criteria was invited to be surveyed.

4.5. Validity and quality checks

4.5.1. Representativeness and external validity

To achieve external validity, except for Norway, we randomly selected the teachers who were surveyed. Although a maximum number of 20 teachers per school were randomly selected, given the difficulties in accessing schools, the instructions given to the surveyors were to ensure a minimum response rate among the randomly selected teachers, which varied by country. The main criterion was to obtain a margin of error of three percent with a 95% confidence interval and, given the survey's multiple purposes, a value of $p = .5$. As can be seen in Table 4, in all three countries we obtained a very low margin of error: two and a half percent in Norway, three point three percent in Spain (Catalonia), and two point eight percent in Chile. In Chile and Norway, we surveyed more teachers per school than the minimum number established. Although in Catalonia we did not reach the minimum number of between 13 and 14 teachers per school, we still managed to survey an average of 11 teachers per school and be just above that percentage (3.3%).

Even with such small error margins, we face certain limitations that we consider when analyzing our results. Since in all the countries not all the 20 selected teachers per school participated in the survey and, in the case of Norway, the selection of teachers was not random, there is risk of selection bias that could affect the sample composition. We cannot address this bias because we do not have information on the individual observable characteristics of the population, making it impossible to compare it with our sample. Nonetheless, it is important to highlight an

Table 4
Teacher sample and margin of error.

	Teacher population (selected urban areas)	Surveyed schools	Teachers to be surveyed to have a 3% margin of error		Surveyed teachers		Margin of error obtained
			Total	Per school	Total	Per school	
Chile	40,000	81	1,040	13	1,185	14.6	2.8%
Catalonia (Spain)	95,000	78	1,056	13.5	852	11.0	3.3%
Norway	20,300	159	1,014	6.4	1,389	8.7	2.5%

inherent aspect of survey experiments. Recent studies have found similarities between results obtained from convenience samples and those based on population samples, reinforcing the validity of using non-probability samples in experimental research (Mullinix et al., 2015). This stems from the nature of the conjoint technique used in our study, which unlike observational studies, ensures the absence of correlations between the attributes of the presented working scenarios and the respondent characteristics. This provides an unbiased estimation of the causal effect of each attribute in explaining the respondents' preferences and enables us to compare the effects of multiple attributes on the same scale (Druckman & Leeper, 2012).

4.5.2. Internal validity

The core principle of experiments is randomization that assures the quality of the achieved experimental data collected and their internal validity, that is, unbiased causal effects' estimations. In conjoint experiments, successful randomization implies the absence of correlation between the attributes displayed in the scenarios, as well as between the attributes and respondent's characteristics. To test whether these two properties of conjoint experiments, known as orthogonality and balance, are fulfilled in our experiment, we calculate bivariate Pearson's correlations (r) between the conjoint attributes and between the conjoint attributes and some key characteristics of the respondents. As shown in Appendix A (Tables 1 and 2), all correlations are all close to zero demonstrating that both principles are met (cf. Campbell & Stanley, 1963). As a result, the estimates are unbiased and internally valid (see Kuhfeld et al., 1994). Additionally, as can be consulted in Appendix A (Table 3), all the conjoint attributes and levels were equally likely to be displayed.

4.6. Empirical models

To analyze the data of the conjoint experiment, collected among 2,924 teachers, we reorganized the dataset so that each scenario, *k*, of task, *j*, presented to a respondent, *i*, is in a different row, so that we generated a total of 11,292 observations.

Whether respondent, *i*, chooses a working scenario, *k*, in task, *j*, is modeled as a function of a vector that contains the attributes of the scenarios proposed to the respondent in that task:

$$y_{ijk} = X_{ijk}\beta + e_{ijk}$$

As respondents answer two tasks, we cluster standard errors by respondents to take into account that answers given by the same respondents are related. As the design does not have restricted profiles, and randomization was uniform across profiles (all combinations of attributes are equally likely to be displayed), the coefficients obtained with a linear regression coincide with the average marginal component effects (AMCEs), i.e. "the average difference in the probability that an option will be preferred when comparing an attribute with its reference category where the average is taken over all possible combinations of the other attributes" (Hainmueller et al., 2014, p. 19). More specifically, we coded the responses to the question on the preferred scenario as a binary variable, namely, 1 if the scenario was chosen and 0 otherwise (dependent variable). We regressed this dependent variable on a series of categorical variables that, for each dimension, took a different value according to the attributes visualized by the respondents. The linear regression thus provides estimates of to what extent a given attribute negatively or positively affected the probability of a given working scenario to be chosen by the surveyed teachers.

To address our first and third research questions, we identify the most significant dimensions for teachers and compare it across the three policy settings. Using the coefficients obtained with the linear regression, the following formula allow us to calculate the preference score for each dimension, i.e. the relative importance of a dimension in explaining the choice compared to the others (Kotri, 2006):

$$S_a = \left(\frac{(max\ u_a\ -\ min\ u_a)}{\sum_{p=1}(max\ u_a\ -\ min\ u_a)} \right)$$

where S_a is the relative importance of a dimension, max u_a is the utility of the dimension's most preferred attribute, and min u_a is the utility of the dimension's least preferred attribute. As can be remarked, preference scores refer to the dimension utility range, i.e. difference between the highest and the lowest coefficients of the attributes of a dimension. The larger the range, the greater the relative importance of a dimension in influencing a teacher's choice of a given working scenario. Preference scores indicate the relative importance of a dimension compared with the other dimensions displayed. This means that they are relative to the other dimensions used in this study and should therefore not be interpreted in an absolute manner.

To compare the importance that teachers assign to each attribute on the same scale (and not in reference to a specific baseline category) and to enable comparison within and across countries (second and third research questions), after having estimated the AMCEs with the linear regression, we calculate the conditional marginal means to compare preferences between different subgroups, i.e. teachers' preferences in the three countries studied. Indeed, Leeper et al. (2020) highlight that comparing the results of the linear regression to check for differences and similarities in preferences of different subgroups can be misleading as the subgroup preferences concerning the reference categories normally differ. Besides, marginal means allow us to visualize and compare regarding the reference categories of each dimension. We also perform F-tests to check for the presence of any significant differences between the results obtained in the three contexts.

5. Results and discussion

In this section, we first report results on the importance of each work dimension to understand what matters most to teachers in the three countries (subsection 5.1). Next, in subsection 5.2, we display the relative importance of each attribute in explaining teachers' choices in relation with each work dimension. In both subsections, we compare the results across the three cases, dimension by dimension, initially outlining the findings of the conjoint experiment. We interpret these results, considering factors such as professional context, school governance, and accountability regulatory frameworks in each country.

One feature that should be considered when interpreting these results is the relative nature of the preferences. When using conjoint experiments, we measure how respondents evaluate one option in comparison to the others. This method allows us to understand how attributes weigh against each other, and determine which attributes are

the most preferred and influential in decisions, but does not indicate absolute preferences. While it captures the order of preferences among the attributes displayed, it does not inform us of the intensity or importance that such preferences have for teachers. Understanding this limitation is crucial to avoid misconceptions when considering the policy implications of our findings.

To improve clarity and facilitate intuitive understanding, we have chosen to graphically present the key findings (preference scores and marginal means) within the main body of the paper. A tabular representation of all results, including overall and country-specific AMCE estimates, can be found in Appendix B.

5.1. Relative importance of the work dimensions by country

To identify which work dimensions, among the ones included in our study, are most significant for teachers, we display the preference scores for each dimension in Fig. 1. As can be seen, the work dimensions that matter the most for teachers are, in this order, goal-setting, assessment of teaching quality, and the existence of outcome-based financial incentives. Yet, there are some differences between countries that merit discussion.

Contrary to Norway and Chile, when choosing their preferred working scenario, teachers in Catalonia (Spain) prioritize how their work is evaluated over the clarity and communication of goals. How goals are defined and communicated emerges as the second most important dimension. This could be related to the fact that in Catalonia the modality of teacher evaluation is predominantly bureaucratic, and new forms of external and performance-based evaluation generate both expectation and concern. While the dimension of teaching quality's assessment is central in Catalonia, it is considered secondary in Chile and Norway. The most important dimension when deciding the preferred working scenario in Norway and Chile is goal-setting, with significant importance also given to the presence or absence of outcome-based financial rewards. The dimension to which teachers in all countries give least importance when choosing a working scenario is the support received from different school actors.

5.2. Teachers' preferences regarding specific attributes by each work dimension

The analyses of teachers' preferences regarding specific policies and practices (or attributes) highlight important differences between the three countries in relation to most work dimensions. The pairwise comparisons (Omnibus F-tests) carried out between the three cases across all the attributes to identify statistically significant differences show that these overall differences are non-zero (Chile-Norway: $F(11, 8549) = 30.58$, $p < 0.001$; Chile-Catalonia: $F(11, 7082) = 9.97$, $p < 0.001$; Norway-Catalonia: $F(11, 6567) = 5.83$, $p < 0.001$). As there are highly significant differences between the results in the three cases, we present the results for each case separately by means of side-by-side graphs (Fig. 2). A single graph with pooled results would average out, and therefore neutralize, the results regarding attributes with opposite effects on the choice.[4]

Fig. 2 contains three side-by-side graphs displaying the marginal mean of each attribute (dot) with 95% confidence intervals (line) for each of the countries under analysis. In each graph, when the dot representing an attribute is located on the right side of the figure, it indicates that this attribute increases the likelihood of the scenario being preferred; when the dot is located on the left side, this means that the attribute discourages teachers from choosing the scenario. Each graph provides rich information on the statistical significance of the effect of each attribute and allows for statistical comparison between the attributes within each country. More specifically, when the line accompanying the dot crosses the 0.5 line, the result is not statistically significant. The overlap between the lines referring to different attributes indicates that there is no statistically significant difference between the importance given to the attributes; while when there is no overlap, this means that the difference is statistically significant.

Next, we present and discuss the results obtained for each dimension separately. Apart from presenting and discussing the results country by country, we also compare the results across countries. To check for any statistically significant difference across countries, for each dimension, we performed F-tests comparing two equations: one estimating only the effects of the attributes, and another measuring the interaction between the case and the attributes. Results of these tests are presented within the text for each dimension.

5.2.1. Dimension #1: goal-setting

In all three countries, teachers prefer to work in schools with well-defined and well-communicated objectives. Although Chilean teachers seem to attribute greater importance to having clear and well defined objectives, the overlap of the confidence intervals indicates that there are no significant differences in relation to this dimension between the three countries. While differences are observed among all countries, the preferences between Norway and the other two countries are clearer and statistically significant at the 1% level (Chile-Norway: $F(3, 8565) = 9.42$, $p < 0.001$; Chile-Catalonia: $F(3, 7098) = 2.91$, $p = 0.033$, Norway-Catalonia: $F(3, 6583) = 3.39$, $p = 0.017$).

In all three contexts, results show that teachers prefer having clear objectives over operating without them or with vague ones. In Chile, a clear hierarchy of preferences is observed according to the definition and communication of the objectives, with a preference for well-defined and well-communicated objectives, a moderate non-preference towards the existence of 'goals even if they are not clear', and a manifest rejection of the absence of goals. A plausible explanation is the long-lasting outcomes-based management regime in Chile, which is an essential part of teacher work and school management. This has led teachers to rely more on goal-setting and strategic planning practices (Ávalos & Bellei, 2019).

In Norway, the rejection of the absence of objectives is less pronounced than in Chile. This could be explained by the greater level of professional autonomy of teachers and a greater capacity to define their own goals (Montecinos et al., 2020). In Catalonia, teachers similarly favor clear objectives, akin to Norway and Chile, but with less intensity. This may partly reflect the value placed on teaching autonomy and horizontality, and the recent, but still underdeveloped implementation of outcomes-based management in schools (Verger & Curran, 2014).

5.2.2. Dimension #2: assessment of teaching quality

The teacher evaluation dimension is where we find more similarities between teachers' preferences in the three countries. This dimension is indeed the only one in relation to which teachers, independently of the country setting, agree and do not show any statistically significant differences (Chile-Norway: $F(3, 8565) = 0.63$, $p = 0.595$; Chile-Catalonia: $F(3, 7098) = 1.37$, $p = 0.251$; Norway-Catalonia: $F(3, 6583) = 2.37$, $p = 0.194$). Teachers in the three countries prefer to be evaluated for their classroom practices through observation and dislike being evaluated by performance according to external standardized tests.

Classroom observation is the most preferred assessment method of teacher quality in the three countries, ahead of standardized tests and the teacher portfolio. Teachers prefer observation as an assessment method, as opposed to external tests that focus on student achievement, or to the teaching portfolio, which is based on the self-evaluation of one's own work. Teachers might welcome classroom observation as they are likely to see it as professionally driven, since it is often associated with peer evaluation. Furthermore, this instrument tends to provide immediate and meaningful feedback to teachers, thus facilitating the improvement of teaching strategies and approaches, especially when it

[4] The tabular presentation of all results, including the pooled results of the linear regression (AMCEs), can be found in Appendix B.

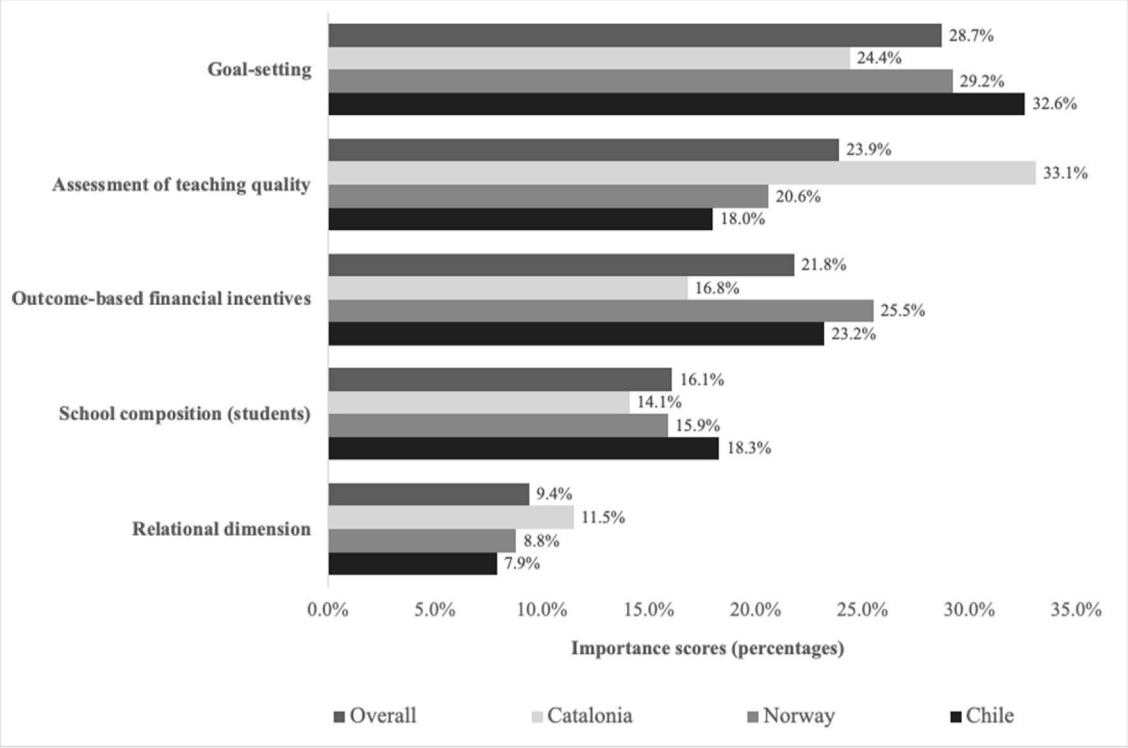


Fig. 1. Preference scores assessing the relative importance of each work dimension.
Source: Reformed database. Catalonia, Norway, and Chile.

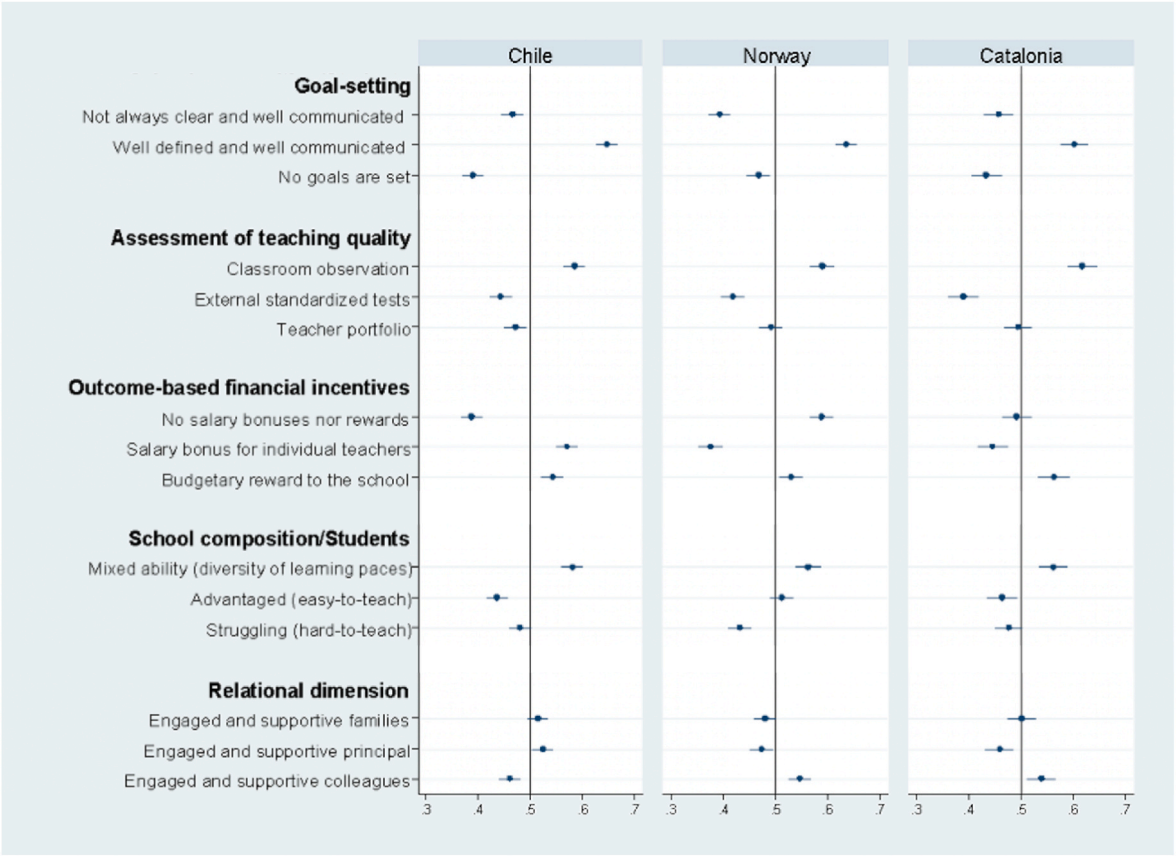


Fig. 2. Teachers' preferences regarding specific attributes (marginal means with 95% CI).
Source: Reformed database. Catalonia, Norway, and Chile.

comes from an experienced colleague (O'Leary, 2022; Range et al., 2013).

In contrast, teachers coincide in rejecting external standardized tests as a teaching quality assessment tool – despite teachers' experience with this type of instrument differs substantially in the three countries. This phenomenon is likely to be linked to the reductionist nature of performance-based assessments, which focus on a narrow part of teachers' work, ignoring the teaching process and privileging learning outcomes (Au, 2007; Mintrop & Sunderman, 2013). It may also have to do with the technical difficulties – when attempted-of separating teaching quality from the influence of schools' social composition on results, reason why these assessments are frequently perceived as unfair measures of teaching quality (Baker et al., 2010).

Finally, the teacher portfolio is a assessment tool that rather generates indifference among teachers in the three countries, even though it is a qualitative, process-assessment strategy that teachers tend to prefer to performance-based assessments, and that has spread among pedagogical circles. Although teachers prefer the teacher portfolio over evaluations based on standardized tests, they do not consider it positively. In Chile, where teacher portfolios are a core component of teacher evaluation, they are perceived as negatively as the standardized tests. This may be because teachers conceive this method as burdensome paperwork (Heneman III et al., 2006; Santiago et al., 2013) with limited effectiveness, especially in the absence of meaningful feedback (Tummons, 2010).

5.2.3. Dimension #3: outcome-based financial incentives

Regarding outcome-based financial incentives, in all three countries there is a preference for collective budgetary rewards (that is, rewards that go to schools, instead of to the individuals). Apart from this element in common, many other differences emerge. In fact, the differences between the countries in this dimension are statistically significant (Chile-Norway: $F(3, 8565) = 72.20$, $p < 0.001$; Chile-Catalonia: $F(3, 7098) = 19.54$, $p < 0.001$; Norway-Catalonia: $F(3, 6583) = 9.92$, $p < 0.001$).

In Norway, teachers prefer not to have any incentive scheme at all, although they prefer budgetary rewards for the school to any type of individual salary bonus. In Chile, the situation is the opposite: teachers prefer both individual salary bonuses and rewards to schools as incentive policies, while expressing a great rejection towards the absence of monetary incentives. Finally, in Catalonia, teachers dislike individual salary bonuses and prefer budgetary rewards that go to the school. However, this result should be interpreted with caution, as this dimension is less determinant in Catalonia than in the other two countries, as indicated by the preference score.

Salary bonuses in Chile, a common and normalized policy, are likely more valued due to lower teacher salaries and satisfaction levels compared to Norway and Catalonia. In Chile, 65% of lower secondary teachers are in schools where appraisals can lead to salary increases, a practice only found in 4% of Norwegian schools (OECD, 2020). Norwegian and Catalan teachers enjoy salaries above the OECD average and higher satisfaction with their salaries, which is something that could explain why they place less emphasis on individual bonuses. Chilean teachers, facing lower salaries even after purchasing power adjustments and less salary satisfaction, might view these bonuses as a necessary supplement to their inadequate income.

The rejection of individual salary bonuses in Norway and Catalonia might also be linked to the characteristics of professional relationships. In contexts of great collegiality among teachers, such as the Norwegian one (OECD, 2020), or horizontal work relations, as in Catalonia, where, as we will show below, there is a greater preference for the support received from colleagues, individual incentives might be perceived as a threat to cohesion and collaboration. This finding would echo several international experiences that indicate that, when incentives are assigned to individual teachers, they have a potentially divisive impact on the school's teaching staff (e.g. Marsden, 2004; Sutherland et al., 2018).

5.2.4. Dimension #4: school composition (students)

The 'School composition' dimension concerns the sociodemographic characteristics of the students in class. The attributes in this dimension refer to working in disadvantaged environments (with so-called *hard-to-teach students*), in advantaged contexts (with, presumably, *easier-to-teach students*), or with mixed-ability students. As Fig. 2 shows, teachers in the three countries have a clear preference for working in environments with mixed-ability students, while preferences in relation to the other attributes show statistically significant differences between Chile and the other two countries. The F-test shows clear statistically significant differences between Chilean and Norwegian teachers ($F(3, 8565) = 8.97$, $p < 0.001$, and between Chileans and Catalans ($F(3, 7098) = 3.63$, $p = 0.012$), while the differences between Norwegians and Catalans are unclear, at the limit of significance of 10% ($F(3, 6583) = 2.37$, $p = 0.068$).

As Fig. 2 displays, the results concerning the attribute 'teaching struggling students', which is usually associated with teachers' prosocial value orientation, are more accentuated in Chile and Catalonia (Spain). In both countries, we observe a non-preference for working in environments with a socially advantaged composition, while in Norway this attribute is unimportant, as it is not statistically significant. In contrast, Norway is the only country where teachers declare a clear, statistically significant, non-preference for working in socially complex environments. This result is consistent with other studies that indicate Norway scores comparatively lower in pro-social values as motivators for teachers entering the profession. According to these studies, Norwegian teachers strongly uphold egalitarian principles in education and often work in socially balanced schools. Consequently, in this country, there is less need to externalize pro-social motives (Watt et al., 2012).

Nonetheless, it is worth noting that, in Catalonia and Chile, the preference towards socially challenging contexts does not appear to be significant. This may be because the Catalan and Chilean education systems are highly segmented, which makes teachers aware of the challenges involved in working in highly vulnerable schools. Such school contexts often have insufficient administrative support, high teacher turnover rates, narrowed pedagogical and organizational processes, and school environments with discipline issues (Glazer, 2021). Hence, not opting for disadvantaged contexts would suggest that preferences based on personal utility implicitly outweigh those centered on prosocial considerations.

Results in Fig. 2 also reflect teacher cross-country preference for working in environments with mixed-ability students. Thus, they opt for heterogeneous school contexts, which are not socially segregated, neither at the top nor at the bottom of the socioeconomic gradient. As a preference, teachers are more inclined to deal with the challenges found in contexts characterized by diversity, but would avoid challenging contexts where hard-to-teach students are concentrated. In Catalonia, the preference for mixed-ability classes aligns with recent discourses advocating for student distribution across schools, as a measure to tackle school segregation and educational inequality (Síndic de Greuges, 2019). Promoting the distribution of socially disadvantaged students across schools has also been a theme of educational debate and reform in Chile in the last decade (Ávalos & Bellei, 2019). As far as Norway is concerned, this preference would be in line with the principle and practice of unity and equality in Norwegian schools, ensuring students of different profiles and learning paces learn together (Lyng & Blichfeldt, 2003).

5.2.5. Dimension #5: relational dimension

Regarding the relational dimension, there are similarities in the preferences expressed by Norwegian and Catalan teachers, while there are differences with respect to Chilean teachers. These differences are highly statistically significant between Norway and Chile, and between Chile and Catalonia, whereas no significant differences appear between Norway and Catalonia (Chile-Norway: $F(3, 8565) = 10.29$, $p < 0.001$; Chile-Catalonia: $F(3, 7098) = 8.37$, $p < 0.001$; Norway-Catalonia: F (3,

6583) = 0.70, $p = 0.526$).

Firstly, we note that receiving support from families is not a statistically significant preference in any of the countries, nor are there significant differences between them. In contrast, there are differences regarding support that comes from the school principal or from colleagues: for Chilean teachers it is more important to have supportive principals, whereas in Norway and Catalonia it is of most importance to have supportive colleagues. Even though support is the least determinant dimension in the three countries, there are reasons to believe that differences in school governance traditions explain these differences.

In Chile, the most plausible explanation is that a model of professionalization of the managerial role of school principals has been installed. This has led to greater acceptance of the leading and hierarchical role of the principal, and to the expected support from this figure to develop a good work (Weinstein & Muñoz, 2012). In contrast, in both Norway and Catalonia, a tradition of democratic and horizontal school governance would explain why peer support is more important than principal support. In Catalonia, principals are often seen as equals (primus inter pares), with their professional roles still evolving despite legislative efforts to distinguish them from teaching roles (Verger & Curran, 2014). Consequently, Catalan teachers view principals as colleagues rather than distinct managerial figures. In Norway, teacher collaboration, particularly in student learning discussions and evaluation standardization, is highly prevalent and integral to the profession, as highlighted in TALIS 2018 (OECD, 2020). This ingrained and taken-for-granted culture of peer support and collaboration may also explain why within-school support ranks lowest in importance among teachers in these three countries.

6. Conclusions

This paper explores teachers' preferences regarding specific work dimensions that have been affected by recent school governance reforms, offering original insights into the policies and practices that teachers may find more motivating and engaging. The paper also examines the complex interplay between teachers' preferences and educational policy settings, professional traditions, and working conditions. One of its strengths is proposing an experimental methodology that overcomes the limitations of traditional survey question-based research. This approach allows the unbiased identification of the relative importance of some work dimensions and attributes in shaping teachers' preferences, while mitigating potential social desirability biases. Another strength of the study lies in its comparative perspective, which underscores the importance of contextualizing teachers' preferences. This allows us to propose several possible explanations and hypotheses that will enrich future research endeavors.

The findings reveal that the value teachers place on specific work aspects and corresponding school policies varies significantly with context, thus encouraging an approach that considers teachers' preferences as shaped by their broader work environment. In Chile, a country with a long-established high-stakes test-based accountability system, teachers prioritize attributes associated with extrinsic motivation and goal-setting. The importance given to outcome-based financial incentives in this country reflects a widely internalized policy norm in a context with teachers' wages below the adjusted international average. In Norway, while we observe a similar hierarchy of importance, the underlying reasons differ significantly from those in Chile. In Norway, despite valuing extrinsic motivators like budgetary rewards to the schools, the major focus is on the absence rather than the presence of outcome-based incentives,. Whereas in Catalonia, the emphasis is not on external financial rewards but on collective incentives for schools. There, as also happens in Norway, teachers prioritize aspects related to teaching quality assessment methods (favoring qualitative and peer-review approaches) and the definition and communication of performance goals.

Teachers in the three countries exhibit significant differences in valuing rewards such as productivity salary bonuses, something that seems to be influenced by the extent of implementation of these measures (Ryan & Deci, 2020), but also by teachers' work conditions. In Chile, individual financial incentives are perceived as a compensation for low salaries and align with the long-standing tradition of outcomes-based management in education. Conversely, in Norway, where collegiality is highly valued, individual bonuses might be seen as potentially threatening positive collegial relationships.

Nonetheless, teachers worldwide also share certain preferences, supporting Letendre (2021)'s notion of transnational commonalities in teaching beliefs. Internationally favored aspects include working in socially mixed classes, employing qualitative and process-oriented teaching assessment methods (like classroom observation), having clear, well-communicated goals, and collective budgetary rewards at the school level. Generally, teachers across different regions concur on the significance of overarching school reform principles like goal-setting and teacher assessment, while commonly rejecting contentious NPM tools that foster competition, hierarchies, and performance pressure.

The practical implications of these findings underscore the importance of adopting a nuanced approach to educational policy that takes into account teacher preferences. Across diverse contexts, our findings invite decision-makers to focus on policies that nurture teacher professionalism and foster collaboration, as well as desegregation policies that promote social diversity in schools. Prioritizing these aspects can bolster teacher motivation and engagement, ultimately creating a more conducive educational environment and improved student outcomes. To optimize the effectiveness of these policies, it is crucial to tailor them to reflect the cultural and policy landscape of each country. For example, in countries like Norway, where intrinsic motivation and collegiality hold significant value, policy designs should lean towards fostering intrinsic rewards, emphasizing collective achievements, and employing qualitative assessment methods. Conversely, in places like Chile, where financial considerations significantly influence teacher engagement, school governance policies need to cater teachers' material needs and expectations. This tailored approach ensures that educational policies are better aligned with the unique policy expectations and preferences of teachers in different settings, thereby enhancing their sense of fulfillment and positively impacting educational settings.

While this study contributes valuable insights, it is not without limitations, necessitating further research. One limitation pertains to the focus on a restricted set of work dimensions and the relative nature of preference scores. Future research may wish to further validate the set of work dimensions and attributes presented in our experiment. Additionally, future studies should explore whether the way in which the attributes are presented may influence the results obtained. Furthermore, the hypotheses and explanations we proposed in this study to explain our findings require further exploration.

Future research should examine teachers' preferences within individual countries to see how school conditions, collegial dynamics, performance pressures, and pedagogical beliefs interact and affect teachers' work preferences. Further studies could also explore relations between teachers' preferences for outcome-based financial incentives and factors like average teacher salaries or a country's GDP. Moreover, the identified preferences could guide subsequent research into the causal relationships between preferred working conditions and key outcomes such as teachers' well-being, engagement, and performance.

To sum up, our findings emphasize the need to consider professional traditions, regulations, general working conditions, and educational policy context for a comprehensive understanding of teachers' work preferences and policy priorities. This research suggests adopting school policies that take into account teachers' ideas aiming at bolstering their motivation and satisfaction. Despite its assumed pro-social, personal, and vocational nature, teacher motivation and engagement can be nurtured and sustained through public policy. This consideration is crucial, as policies resonating with teachers not only foster their well-being and elevate morale, but also enhance excellent and engaging

teaching.

Funding

This work was supported by the European Research Council under Grant 680172 (Reformed Project).

CRediT authorship contribution statement

Antonina Levatino: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Gerard Ferrer-Esteban:** Writing – review & editing, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Antoni Verger:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors have no relevant financial or non-financial interests to disclose or competing interests to declare that are relevant to the content of the article.

Data availability

Data will be made available on request.

Appendix A. Experiment quality checks

Table A1
Correlations between conjoint attributes.

	1	2	3	4	5
1. School composition/Students	1.0000				
2. Assessment of teaching quality	−0.0093	1.0000			
3. Goal-setting	0.0155	0.0113	1.0000		
4. Relations/Support	−0.0023	−0.0144	0.0023	1.0000	
5. Outcome-based financial rewards	−0.0067	−0.0030	−0.0038	0.0074	1.0000

Source: Reformed database. Catalonia, Norway, and Chile.

Table A2
Correlations between conjoint attributes and respondents' characteristics.

	Case	School where respondent is working	Type of provider	Gender
School composition/Students	0.0160	0.0100	−0.0029	−0.0105
Assessment of teaching quality	−0.0008	−0.0095	0.0173	−0.0068
Goal-setting	−0.0051	−0.0055	0.0003	−0.0083
Relations/Support	0.0064	0.0074	−0.0088	−0.0100
Outcome-based financial rewards	−0.0024	−0.0041	−0.0086	−0.0105

Source: Reformed database. Catalonia, Norway, and Chile.

Table A3
Frequencies of conjoint attributes and levels.

ATTRIBUTES AND LEVELS	N	%
School composition/Students		
Mixed ability (diversity of learning paces)	3687	32.65
Advantaged (easy-to-teach)	3766	33.35
Struggling (hard-to-teach)	3839	34.00
Assessment of teaching quality		
Classroom observation	3774	33.42
External standardized tests	3814	33.78
Teacher portfolio	3704	32.80
Goal-setting		
Not always clear and well-communicated	3705	32.81
Well defined and well-communicated	3811	33.75
No goals are set	3776	33.44
Relations/Support		
From families	3818	33.81
From the principal	3695	32.72
From colleagues	3779	33.47
Outcome based financial rewards		
No salary bonuses nor rewards	3867	34.25
Salary bonus for individual teacher	3679	32.58
Budgetary reward to the school	3746	33.17
Total *N* tables	**11,292**	**100**

Source: Reformed database. Catalonia, Norway, and Chile.

Appendix B. Tabular estimates

Table B1
Average marginal component effect (AMCE) estimates.

	Pooled results	Catalonia	Chile	Norway
School composition/Students				
Mixed ability (diversity of learning paces)	Ref.	Ref.	Ref.	Ref.
Advantaged (easy-to-teach)	0.099*** (0.01)	0.098*** (0.02)	−0.144*** (0.02)	−0.052** (0.02)
Struggling (hard-to-teach)	−0.105*** (0.01)	−0.085*** (0.02)	−0.100*** (0.02)	−0.132*** (0.02)
Assessment of teaching quality				
Classroom observation	Ref.	Ref.	Ref.	Ref.
External standardized tests	−0.166*** (0.01)	−0.228*** (0.02)	−0.141*** (0.02)	−0.171*** (0.02)
Teacher portfolio	−0.109*** (0.01)	−0.123*** (0.02)	−0.113*** (0.02)	−0.098*** (0.02)
Goal-setting				
Not always clear and well-communicated	Ref.	Ref.	Ref.	Ref.
Well-defined and well-communicated	0.193*** (0.01)	0.144*** (0.02)	0.181*** (0.02)	0.242*** (0.02)
No goal set	−0.009 (0.01)	−0.024 (0.02)	−0.075*** (0.02)	0.075*** (0.02)
Relations/Support				
From families	Ref.	Ref.	Ref.	Ref.
From the principal	0.009 (0.01)	0.043 (0.02)	0.009 (0.02)	−0.006 (0.02)
From colleagues	0.012 (0.01)	0.036 (0.02)	−0.054** (0.02)	0.067*** (0.02)
Outcome-based financial incentives				
No salary bonuses nor rewards	Ref.	Ref.	Ref.	Ref.
Salary bonus for individual teacher	−0.014 (0.01)	−0.046 (0.02)	0.182*** (0.02)	−0.212*** (0.02)
Budgetary reward to the school	0.057** (0.01)	0.070** (0.02)	0.154*** (0.02)	−0.058** (0.02)
Fixed effects	Case & School	School	School	School
N. observations	11292	2602	4584	4106

Source: Reformed database. Catalonia, Norway, and Chile.

Table B2
Conditional marginal means.*

	Catalonia		Chile		Norway	
	est.	SE	est.	SE	est.	SE
School composition/Students						
Mixed ability (diversity of learning paces)	0.562	0.01	0.581	0.01	0.563	0.01
Advantaged (easy-to-teach)	0.464	0.01	0.437	0.01	0.512	0.01
Struggling (hard-to-teach)	0.477	0.01	0.481	0.01	0.432	0.01
Assessment of teaching quality						
Classroom observation	0.618	0.01	0.585	0.01	0.589	0.01
External standardized tests	0.390	0.01	0.444	0.01	0.419	0.01
Teacher portfolio	0.495	0.01	0.472	0.01	0.491	0.01
Goal-setting						
Not always clear and well-communicated	0.458	0.01	0.465	0.01	0.393	0.01
Well-defined and well-communicated	0.603	0.01	0.646	0.01	0.656	0.01
No goal set	0.434	0.01	0.390	0.01	0.468	0.01
Relations/Support						
From families	0.502	0.01	0.515	0.01	0.479	0.01
From the principal	0.459	0.01	0.524	0.01	0.473	0.01
From colleagues	0.539	0.01	0.461	0.01	0.546	0.01
Outcome-based financial incentives						
No salary bonuses nor rewards	0.493	0.01	0.388	0.01	0.588	0.01
Salary bonus for individual teacher	0.447	0.01	0.570	0.01	0.377	0.01
Budgetary reward to the school	0.563	0.02	0.543	0.01	0.530	0.01
School fixed effects	YES		YES		YES	
N. observations	2,602		4,584		4,106	

*Results displayed graphically in Fig. 2 of the manuscript.
Source: Reformed database. Catalonia, Norway, and Chile.

References

Ávalos, B., & Bellei, C. (2019). Recent education reforms in Chile: How much of a departure from market and New Public Management systems? In C. Ornelas (Ed.), *Politics of education in Latin America* (pp. 43–71). Brill. https://doi.org/10.1163/9789004413375_003.

Admiral, W., & Kittelsen Røberg, K. (2023). Teachers' job demands, resources and their job satisfaction: Satisfaction with school, career choice and teaching profession of teachers in different career stages. *Teaching and Teacher Education, 125*, 1–10. https://doi.org/10.1016/j.tate.2023.104063

Ainley, J., & Carstens, R. (2018). *Teaching and Learning International Survey (TALIS) 2018 Conceptual Framework*. OECD Publishing. https://doi.org/10.1787/799337c2-en (OECD Education Working Papers, No. 187, pp. 1–108).

Altrichter, H., Heinrich, M., & Soukup-Altrichter, K. (2014). School decentralization as a process of differentiation, hierarchization and selection. *Journal of Education Policy, 29*(5), 675–699. https://doi.org/10.1080/02680939.2013.873954

Amrein-Beardsley, A. (2014). *Rethinking value-added models in education: Critical perspectives on tests and assessment-based accountability*. Routledge. https://doi.org/10.4324/9780203409909

Anthony, G., & Ord, K. (2008). Change-of-career secondary teachers: Motivations, expectations and intentions. *Asia-Pacific Journal of Teacher Education, 36*(4), 359–376. https://doi.org/10.1080/13598660802395865

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5), 258–267. https://doi.org/10.3102/0013189X07306523

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., Rothstein, R., Shavelson, R. J., & Shepard, L. A. (2010). Problems with

the use of student test scores to evaluate teachers. EPI Briefing Paper No. 278 https://files.epi.org/page/-/pdf/bp278.pdf.
Ball, S. J., & Maroy, C. (2009). School's logics of action as mediation and compromise between internal dynamics and external constraints and pressures. *Compare: A Journal of Comparative and International Education, 39*(1), 99–112. https://doi.org/10.1080/03057920701825544
Bansak, K., Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2018). The number of choice tasks and survey satisficing in conjoint experiments. *Political Analysis, 26*(1), 112–119. https://doi.org/10.1017/pan.2017.40
Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal, 42*(2), 231–268. https://doi.org/10.3102/00028312042002231
Braun, A., & Maguire, M. (2020). Doing without believing – enacting policy in the English primary school. *Critical Studies in Education, 61*(4), 433–447. https://doi.org/10.1080/17508487.2018.1500384
Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand McNally & Company.
Collins, C. (2014). Houston, we have a problem: Teachers find no value in the SAS Education Value-Added Assessment System (EVAAS®). *Education Policy Analysis Archives, 22*(98), 1–42. https://doi.org/10.14507/epaa.v22.1594
Cuevas, R., Ntoumanis, N., Fernandez-Bustos, J. G., & Bartholomew, K. (2018). Does teacher evaluation based on student performance predict motivation, well-being, and ill-being? *Journal of School Psychology, 68*, 154–162. https://doi.org/10.1016/j.jsp.2018.03.005
De Clercq, M., Watt, H. M. G., & Richardson, P. W. (2022). Profiles of teachers' striving and wellbeing: Evolution and relations with context factors, retention, and professional engagement. *Journal of Educational Psychology, 114*(3), 637–655. https://doi.org/10.1037/edu0000702
Druckman, J. N., & Leeper, T. J. (2012). Learning more from political communication experiments: Pretreatment and its effects. *American Journal of Political Science, 56*(4), 875–896. https://doi.org/10.1111/j.1540-5907.2012.00582.x
Ehren, M. C., Gustafsson, J. E., Altrichter, H., Skedsmo, G., Kemethofer, D., & Huber, S. G. (2015). Comparing effects and side effects of different school inspection systems across Europe. *Comparative Education, 51*(3), 375–400. https://doi.org/10.1080/03050068.2015.1045769
Ferrer-Esteban, G. (2023). *Sampling strategy (REFORMED Methodological Papers, No. 4)*. REFORMED.
Fitzgerald, T. (2009). The tyranny of bureaucracy: Continuing challenges of leading and managing from the middle. *Educational Management Administration & Leadership, 37*(1), 51–65. https://doi.org/10.1177/1741143208098164
Frey, B. S., & Jegen, R. (2002). Motivation crowding theory. *Journal of Economic Surveys, 15*(5), 589–611. https://doi.org/10.1111/1467-6419.00150
Gallego, A., & Marx, P. (2017). Multi-dimensional preferences for labour market reforms: A conjoint experiment. *Journal of European Public Policy, 24*(7), 1027–1047. https://doi.org/10.1080/13501763.2016.1170191
Gewirtz, S., Maguire, M., Neumann, E., & Towers, E. (2021). What's wrong with 'deliverology'? Performance measurement, accountability and quality improvement in English secondary education. *Journal of Education Policy, 36*(4), 504–529. https://doi.org/10.1080/02680939.2019.1706103
Glazer, J. (2021). The well-worn path: Learning from teachers who moved from hard-to-staff to easy-to-staff schools. *Teaching and Teacher Education, 105*. https://doi.org/10.1016/j.tate.2021.103399
Green, P. E., Krieger, A. M., & Wind, Y. (2001). Thirty years of conjoint analysis: Reflections and prospects. *Interfaces, 31*(3), S56–S73. https://www.jstor.org/stable/25062702.
Grant, A. M. (2007). Relational job design and the motivation to make a prosocial difference. *Academy of Management Review, 32*(2), 393–417. https://doi.org/10.5465/amr.2007.24351328
Green, P., & Rao, V. (1971). Conjoint measurement for quantifying judgemental data. *Journal of Marketing Research, 8*(3), 355–363. https://doi.org/10.2307/3149575
Gunter, H. M., Grimaldi, E., Hall, D., & Serpieri, R. (Eds.). (2016). *New public management and the reform of education: European lessons for policy and practice*. Routledge.
Högberg, B., & Lindgren, J. (2021). Outcome-based accountability regimes in OECD countries: A global policy model? *Comparative Education, 57*(3), 301–321. https://doi.org/10.1080/03050068.2020.1849614
Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences, 112*(8), 2395–2400. https://doi.org/10.1073/pnas.1416587112
Hainmueller, J., & Hopkins, D. J. (2015). The hidden American immigration consensus: A conjoint analysis of attitudes toward immigrants. *American Journal of Political Science, 59*(3), 529–548. https://doi.org/10.1111/ajps.12138
Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preferences experiments. *Political Analysis, 22*(1), 1–30. https://doi.org/10.1093/pan/mpt024
Hansen, J. R., & Jacobsen, C. B. (2016). Changing strategy processes and strategy content in public sector organizations? A longitudinal case study of npm reforms' influence on strategic management. *British Journal of Management, 27*(2), 373–389. https://doi.org/10.1111/1467-8551.12157
Harackiewicz, J. M., & Sansone, C. (2000). Rewarding competence: The importance of goals in the study of intrinsic motivation. In C. Sansone, & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 79–103). Academic Press. https://doi.org/10.1016/B978-012619070-0/50026-X.
Harmsen, R., Helms-Lorenz, M., Maulana, R., & van Veen, K. (2018). The relationship between beginning teachers' stress causes, stress responses, teaching behaviour and attrition. *Teachers and Teaching, 24*(6), 626–643. https://doi.org/10.1080/13540602.2018.1465404
Heneman III, H. G., Milanowski, A., Kimball, S., & Odden, A. (2006). *Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay (CPRE Policy Briefs, RB-45)*. https://doi.org/10.12698/cpre.2013.rb45
Horng, E. L. (2009). Teacher tradeoffs: Disentangling teachers' preferences for working conditions and student demographics. *American Educational Research Journal, 46*(3), 690–717. https://doi.org/10.3102/0002831208329599
Huber, J., & Zwerina, K. (1996). The importance of utility balance in efficient choice designs. *Journal of Marketing Research, 33*(3), 307–317. https://doi.org/10.1177/002224379603300305
Ingersoll, R. M., & May, H. (2012). The magnitude, destinations, and determinants of mathematics and science teacher turnover. *Educational Evaluation and Policy Analysis, 34*(4), 435–464. https://doi.org/10.3102/0162373712454326
Jabbar, H. (2015). "Every kid is money" market-like competition and school leader strategies in New Orleans. *Educational Evaluation and Policy Analysis, 37*(4), 638–659. https://doi.org/10.3102/0162373715577447
Johnston, A. C. (2020). *Teacher preferences, working conditions, and compensation structure (IZA DP No. 13121)*. https://docs.iza.org/dp13121.pdf.
Kanfer, R., Frese, M., & Johnson, R. E. (2017). Motivation related to work: A century of progress. *Journal of Applied Psychology, 102*(3), 338–355. https://doi.org/10.1037/apl0000133
Kotri, A. (2006). *Analyzing customer value using conjoint analysis: The example of a packaging company*. https://doi.org/10.2139/ssrn.950497 (Tartu Working Paper No. 567).
Kraft, M. A., & Lyon, M. A. (2022). *The rise and fall of the teaching profession: Prestige, interest, preparation, and satisfaction over the last half century*. https://doi.org/10.26300/7b1a-vk92. EdWorkingPapers, No. 22-679).
Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*, 537–567. https://doi.org/10.1146/annurev.psych.50.1.537
Kuhfeld, W. F., Tobias, R. D., & Garratt, M. (1994). Efficient experimental design with marketing research applications. *Journal of Marketing Research, 31*(4), 545–557. https://doi.org/10.1177/002224379403100408
Lapuente, V., & Van de Walle, S. (2020). The effects of new public management on the quality of public services. *Governance, 33*(3), 461–475. https://doi.org/10.1111/gove.12502
Latham, G. P., & Locke, E. A. (2007). New developments in and directions for goal-setting research. *European Psychologist, 12*(4), 290–300. https://doi.org/10.1027/1016-9040.12.4.290
Latham, G. P., & Pinder, C. C. (2005). Work motivation theory and research at the dawn of the twenty-first century. *Annual Review of Psychology, 56*, 485–516. https://doi.org/10.1146/annurev.psych.55.090902.142105
Lazarides, R., & Schiefele, U. (2021). Teacher motivation: Implications for instruction and learning. Introduction to the special issue. *Learning and Instruction, 76*, 1–3. https://doi.org/10.1016/j.learninstruc.2021.101543
Leeper, T., Hobolt, S., & Tilley, J. (2020). Measuring subgroup preferences in conjoint experiments. *Political Analysis, 28*(2), 207–221. https://doi.org/10.1017/pan.2019.30
Letendre, G. (2021). Teachers in neoinstitutional and world culture theory. *Comparative Education Review, 65*(4), 750–769. https://doi.org/10.1086/716520
Levatino, A. (2021). *Surveying principals and teachers: Methodological insights into the design of the REFORMED questionnaires* (Methodological Notes No. 2). REFORMED. https://doi.org/10.5281/zenodo.4450773.
Levatino, A., Parcerisa, L., & Verger, A. (2024). Understanding the stakes: The influence of accountability policy options on teachers' responses. *Educational Policy, 38*(1), 31–60. https://doi.org/10.1177/08959048221142048
Levy, A., DeLeon, I. G., Martinez, C. K., Fernandez, N., Gage, N. A., Sigurdsson, S.Ó., & Frank-Crawford, M. A. (2017). A quantitative review of overjustification effects in persons with intellectual and developmental disabilities. *Journal of Applied Behavior Analysis, 50*(2), 206–221. https://doi.org/10.1002/jaba.359
Lingard, B., Martino, W., Rezai-Rashti, G., & Sellar, S. (2015). *Globalizing educational accountabilities*. Routledge. https://doi.org/10.4324/9781315885131
Lortie, D. C. (1975). *Schoolteacher: A sociological study*. University of Chicago Press.
Lyng, S.T., & Blichfeldt, J.F. (2003). Attracting, developing and retaining effective teachers (Country Background Report for Norway, pp. 1-82). OECD. https://www.oecd.org/education/school/2635707.pdf.
Lovison, V. S., & Mo, C. H. (2022). *Investing in the teacher workforce: Experimental evidence on teachers' preferences (EdWorkingPapers, No. 22-528)*. https://doi.org/10.26300/fygj-e132
Manuel, J., & Hughes, J. (2006). 'It has always been my dream': Exploring pre-service teachers' motivations for choosing to teach. *Teacher Development, 10*(1), 5–24. https://doi.org/10.1080/13664530600587311
Maroy, S. T., & Voisin, J. F. (2017). *Think piece on accountability (background paper prepared for the 2017/18 global education monitoring report)*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000259572.locale=en.
Marsden, D. (2004). The role of performance-related pay in renegotiating the "effort bargain": The case of the British Public Service. *Industrial and Labor Relations Review, 57*(3), 350–370. https://doi.org/10.1177/001979390405700302
Mintrop, H. (2004). *Schools on probation: How accountability works (and doesn't work)*. Teachers College Press.
Mintrop, R., & Ordenes, M. (2017). Teacher work motivation in the era of extrinsic incentives: Performance goals and pro-social commitments in the service of equity. *Education Policy Analysis Archives, 25*(44), 1–40. https://doi.org/10.14507/epaa.25.2482
Mintrop, H., & Sunderman, G. L. (2013). The paradoxes of data-driven school reform: Learning from two generations of centralized accountability systems in the United

States. In D. Anagnostopoulos, S. A. Rutledge, & R. Jacobsen (Eds.), *The Infrastructure of accountability: Data use and the transformation of American education* (pp. 23–39). Harvard Education Press.

Montecinos, C., González, Á., & Ehren, M. (2020). From hierarchy and market to hierarchy and network governance in Chile: Enhancing accountability, capacity and trust in public education. In M. Ehren, & J. Baxter (Eds.), *Trust, accountability and capacity in education system reform* (pp. 201–221). Routledge. https://doi.org/10.4324/9780429344855.

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science, 2*(2), 109–138. https://doi.org/10.1017/XPS.2015.19

O'Leary, M. (2022). Rethinking teachers' professional learning through unseen observation. *Professional Development in Education*. https://doi.org/10.1080/19415257.2022.2125551

OECD. (2016). *PISA 2015 results (volume II): Policies and practices for successful schools*. Paris: OECD.

OECD. (2020). *TALIS 2018 results (volume II): Teachers and school leaders as valued professionals*. OECD Publishing. https://doi.org/10.1787/19cf08df-en

Penninckx, M., Vanhoof, J., De Maeyer, S., & Van Petegem, P. (2016). Enquiry into the side effects of school inspection in a 'low-stakes' inspection context. *Research Papers in Education, 31*(4), 462–482. https://doi.org/10.1080/02671522.2015.1076886

Perryman, J., Ball, S., Maguire, M., & Braun, A. (2011). Life in the pressure cooker–School league tables and English and mathematics teachers' responses to accountability in a results-driven era. *British Journal of Educational Studies, 59*(2), 179–195. https://doi.org/10.1080/00071005.2011.578568

Pinder, C. C. (2008). *Work motivation in organizational behavior* (2nd ed.). Psychology Press. https://doi.org/10.4324/9781315734606

Räsänen, K., Pietarinen, J., Pyhältö, K., Soini, T., & Väisänen, P. (2020). Why leave the teaching profession? A longitudinal approach to the prevalence and persistence of teacher turnover intentions. *Social Psychology of Education: International Journal, 23*(4), 837–859. https://doi.org/10.1007/s11218-020-09567-x

Range, B. G., Young, S., & Hvidston, D. (2013). Teacher perceptions about observation conferences: What do teachers think about their formative supervision in one US school district? *School Leadership & Management, 33*(1), 61–77. https://doi.org/10.1080/13632434.2012.724670

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. The Guilford Press.

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology, 61*, Article 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Saban, A. (2003). A Turkish profile of prospective elementary school teachers and their views of teaching. *Teaching and Teacher Education, 19*(8), 829–846. https://doi.org/10.1016/j.tate.2003.03.004

Salokangas, M., & Wermke, W. (2020). Unpacking autonomy for empirical comparative investigation. *Oxford Review of Education, 46*(5), 563–581.

Santiago, P., Benavides, F., Danielson, C., Goe, L., & Nusche, D. (2013). Teacher evaluation in Chile 2013 (OECD reviews of evaluation and assessment in education). *OECD*. https://doi.org/10.1787/9789264172616-en

Saris, W. E., & Gallhofer, I. N. (2014). *Design, evaluation, and analysis of questionnaires for survey research*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470165195

Shoshani, A., & Eldor, L. (2016). The informal learning of teachers: Learning climate, job satisfaction and teachers' and students' motivation and well-being. *International Journal of Educational Research, 79*, 52–63. https://doi.org/10.1016/j.ijer.2016.06.007

Síndic de Greuges. (2019). Pacte contra la Segregació Escolar a Catalunya: Un compromís per a l'èxit educatiu (pp. 1–55). Síndic de Greuges. https://www.sindic.cat/site/unitFiles/6058/Pacte%20segregacio%20escolar%202018_definitiu.pdf.

Sullivan, A., Johnson, B., Simons, M., & Tippett, N. (2021). When performativity meets agency: How early career teachers struggle to reconcile competing agendas to become 'quality' teachers. *Teachers and Teaching, 27*(5), 388–403. https://doi.org/10.1080/13540602.2020.1806050

Supovitz, J. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change, 10*(2), 211–227. https://doi.org/10.1007/s10833-009-9105-2

Sutherland, A., Prideaux, R., Belanger, J., Broeks, M., Shenderovich, Y., & Van der Staaji, S. (2018). Motivating teachers with incentivised pay and coaching randomised control trial (ICR trial). *Closure report: Understanding factors influencing participant recruitment failure (research report, No. RR-2145-EEF)*. RAND Corporation. https://doi.org/10.7249/RR2145

Tummons, J. (2010). The assessment of lesson plans in teacher education: A case study in assessment validity and reliability. *Assessment & Evaluation in Higher Education, 35*(7), 847–857. https://doi.org/10.1080/02602930903125256

Verger, A., & Curran, M. (2014). New public management as a global education policy: its adoption and re-contextualization in a Southern European setting. *Critical Studies in Education, 55*(3), 253–271.

Viano, S., Pham, L. D., Henry, G. T., Kho, A., & Zimmer, R. (2021). What teachers want: School factors predicting teachers' decisions to work in low-performing schools. *American Educational Research Journal, 58*(1), 201–233. https://doi.org/10.3102/0002831220930199

Voisin, A., & Dumay, X. (2020). How do educational systems regulate the teaching profession and teachers' work? A typological approach to institutional foundations and models of regulation. *Teaching and Teacher Education, 96*, 103–144.

Watt, H. M. G., Richardson, P. W., Klusmann, U., Kunter, M., Beyer, B., Trautwein, U., & Baumert, J. (2012). Motivations for choosing teaching as a career: An international comparison using the FIT-Choice scale. *Teaching and Teacher Education, 28*(6), 791–805. https://doi.org/10.1016/j.tate.2012.03.003

Weinstein, J., & Muñoz, G. (2012). *¿Qué sabemos sobre los directores de escuela en Chile?* Fundación Chile & CEPPE.

World Bank. (2024). GDP per capita, PPP (current international $). Retrieved from https://data.worldbank.org.

Zhao, Y. (2017). What works may hurt: Side effects in education. *Journal of Educational Change, 18*, 1–19. https://doi.org/10.1007/s10833-016-9294-4