



Standardised testing in the context of constitutionally protected freedom of education – the case of Flanders

Michelle Meadows & Inés Sanguino

To cite this article: Michelle Meadows & Inés Sanguino (2025) Standardised testing in the context of constitutionally protected freedom of education – the case of Flanders, Oxford Review of Education, 51:2, 202-222, DOI: [10.1080/03054985.2024.2439283](https://doi.org/10.1080/03054985.2024.2439283)

To link to this article: <https://doi.org/10.1080/03054985.2024.2439283>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 14 Mar 2025.



Submit your article to this journal [↗](#)



Article views: 31



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)



Standardised testing in the context of constitutionally protected freedom of education – the case of Flanders

Michelle Meadows ^a and Inés Sanguino

^aDepartment of Education, Oxford University, Oxford, UK

ABSTRACT

Around the world governments have been observed to harness assessment to exert control over the content and quality of education. In Belgium, Flanders, with constitutionally protected freedom of education, has been an exception, having no standardised tests or end of school examinations. Recently, declining international test outcomes have been used to justify educational reforms, including the introduction of compulsory standardised tests. We consider the timing and rationale for the reforms, and the extent to which the changes fit the Global Education Reform Movement template. We conclude that the policy reforms represent an attempt by government to leverage Broadfoot's social functions of assessment to create a different balance between school autonomy and accountability. How the reforms will interact with the educational culture in Flanders is difficult to predict but the new quantification of education will likely open the system up to further policy reforms as test data driven comparisons raise questions and demand action.

KEYWORDS

Flanders; standardised tests; policy; accountability

Introduction

National standardised tests have become a feature of education systems in many countries around the world, coming to dominate teaching and learning. Flanders (the Flemish Community of Belgium) has been a notable exception but in a recent series of education policy reforms, compulsory standardised tests are being introduced in primary and secondary schooling. In this paper, utilising publicly available policy documents and research papers, and drawing on an Organisation for Economic Co-operation and Development (OECD) review conducted in part by the authors (OECD, 2023c), we employ Flanders as a case study to explore government policy-making in a decentralised system, where freedom of education is legally protected and tools for central control have been lacking. Flanders is an outlier in coming late to compulsory standardised testing – making it a worthwhile case study for its uniqueness (Yin, 2009). We consider how the policy changes compare to the typical pattern of recent global education reforms (Sahlberg, 2016, 2023), including in harnessing two powerful social functions of assessment identified by Broadfoot (2021) – the use of assessment as a means of control of the quality of

CONTACT Michelle Meadows  michelle.meadows@education.ox.ac.uk  Department of Education, Oxford University, 15 Norham Gardens, Oxford OX2 6PY, UK

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

education and to determine teaching and learning via the content of what is taught. We reflect on the likely consequences of the changes for education in Flanders.

Social functions of assessment

Large-scale assessments are more than technical tools to measure educational attainment; they perform significant social functions too. Broadfoot (2021) argued that these social functions mediate the relationship between education and society. She defined the social functions of assessment in terms of their historical emergence: first, to attest competence, a function rooted in the history of craftsmanship and found in vocational and technical education; second, the use of assessment as a rational means of regulating competition for limited opportunities and resources; third, the role of assessment in determining the content of what is taught and, accordingly, the priorities of teachers; fourth, the use of assessment as a means to control the quality of education, for example through mechanisms of accountability. The social functions of assessment render testing a potent policy lever for governments around the world. We argue that in the case of Flanders, in a changing political and social context, they are being leveraged by government to exact a greater degree of central control over education.

Assessment as a policy lever

National large-scale assessments allow governments to ‘steer at a distance’ the complex network of providers and actors that make up any school system (Ball, 1994; Broadfoot, 2021). They are the tool by which governments ensure that those working within schools meet targets and follow the teaching and learning standards defined in the curriculum (Ball, 1994; Verger, Parcerisa, et al., 2019). As Broadfoot (2021) put it ‘... a centralized, government-controlled, external assessment apparatus possesses great power to enforce the pursuit of acentrally determined curriculum’ (p. 40). It allows governments to avoid prescriptive control of education and to govern by incentives rather than constraints, giving the appearance of autonomy for those working in the system (Ball, 1994). Rather than allowing autonomy, however, the assessment content, format and marking criteria determine what students and teachers need to do to achieve ‘success’ (Godfrey-Faussett & Baird, 2025).

Teaching to the test is unavoidable to the extent that teachers want to improve students’ test scores (Alderson & Wall, 1993). Improvements are incentivised by school accountability systems focused on students’ outcomes – ‘test-based accountability’ (Hamilton et al., 2002). In test-based accountability outcomes have consequences, for example for teachers’ jobs and for school funding. Test results may be published and used by parents and carers to inform school choice, positioning parents as consumers of education and schools responsive to competitive pressures (Olmedo & Wilkins, 2017). Given the stakes, the tests come to dominate education (Russell et al., 2009).

This global shift in the uses of assessment has been seen as key part of a neoliberal Global Education Reform Movement (GERM). School competition and parental choice, aligning teaching and learning to common standards, standardised testing, particularly in literacy and numeracy, using test data to hold schools to account, and new forms of public-private partnership have been the main policy principles of the GERM (Sahlberg,

2016, 2023). According to Verger, Parcerisa, et al. (2019), GERM reflects the fact that in recent decades, most educational reforms around the world have responded to similar problems and priorities, and have followed a similar policy rationale. Although Anglo-Saxon and countries dependent on aid first embraced the GERM, the approach spread to many other countries, including European and OECD countries. Sahlberg (2016) noted, however, that not all education systems have adopted reforms consistent with the GERM and some expressed reservations about the approach, including Belgium.

The political context in Flanders

Belgium is a federal state with three regions (Flanders, Wallonia, and Brussels), and three cultural communities (Dutch speaking, French speaking, and German speaking), each with their own parliament and government. The Flemish have, however, all but merged their region and community into a single institutional framework (Andeweg, 2019).

Belgium has been considered a classic consociational democracy – a divided society in which conflict is managed by political accommodation at the elite level (Lijphart, 1977). Divisions once occurred along class and religious cleavages but now exist along language lines, and consociationalism is used at intervals to deal with tensions (Andeweg, 2019). However, consociationalism has never been completely accepted in Belgium and politics has remained competitive with a regional focus.

In Flanders, since the 1980s there has been a significant decline in the popularity of the social democratic party (Gaasendam et al., 2021) and the right-wing has increased in power. The conservative New Flemish Alliance (N-VA) party has the largest share of the vote in both Belgium and Flanders. The radical-right Vlaams Belang (VB) party comes a close second. At the time of writing the federal government had yet to be formed following the June 2024 elections. After the May 2019 elections it took until September 2020 to form a new government, when, with some difficulty, a majority coalition that excluded these nationalist parties was assembled. Naturally, this act was strongly criticised by the N-VA and VB.

While both parties are nationalist, they have distinct approaches. VB has argued for Belgium to be split in the coming years, promising a ‘declaration of sovereignty’, whereas N-VA has positioned itself as moderate, democratic and rational (Maly, 2016), rejecting short-term separatism and advocating allocating minimal powers to the federal government (Maddens et al., 2020). Beyond nationalism, N-VA’s approach is neoliberal – advocating economic liberalism, state retrenchment and immediate tax reductions to stimulate the economy. This extends to education – the Education Minister Ben Weyts is reported as describing students as ‘capital’ and teachers as ‘asset managers’ (Tourne et al., 2024).

While individual responsibility is a crucial part of N-VA’s discourse, nationalism is positioned as a positive force. N-VA presents its nation-building project as based on a civic conception of national identity and different from the ‘xenophobic ethnic nationalism’ propagated by the VB (Gaasendam et al., 2021). Indeed, immigration receives considerable media and political attention in Flanders. There have been high numbers of incoming refugees, asylum seekers, and transit migrants (De Coninck & Joris, 2021). N-VA has supported increasingly assimilationist and exclusionary policies, linked to its nationalist agenda (Adam & Xhardez, 2024). These are

wide-ranging but include plans to ban speaking languages other than Dutch in schools, including outside of the classroom (Belga News Agency, 2023, August 30) and for minimum targets for Dutch-language skills for five-year olds (Chini, 2023a, August 30).

Education in Flanders

Constitutional freedom of education

The three communities of Belgium have their own independent education systems, with few federal level regulations. Even prior to the ascent of the N-VA and neoliberalism, Flanders' system exhibited market-like characteristics due to its strong reliance on the constitutional principle of educational freedom (Tourne et al., 2024). This principle dates back to 1830 (De Groof et al., 2010) and guarantees that anyone can set up a school and decide its policies and teaching practices. Parents are free to enrol their children at their preferred school as long as the school is recognised by the government. For this, schools must meet a set of minimum criteria, which almost all schools do (Shewbridge & Köster, 2021). Funding is allocated to schools on a per student basis. There is a large choice of schools and competition between schools for students is high (Shewbridge & Köster, 2021).

The system is highly decentralised and characterised by a very high degree of school autonomy (OECD, 2018). Responsibility for education quality lies mainly with schools and teachers (Janssens et al., 2016). Government does not regulate educational processes and methods. Schools develop their own policies, pedagogical plans, teaching methods, curriculum, and so on.

Schools also have great autonomy in their assessment practices, especially when compared to many other countries (Shewbridge et al., 2019). Each school's regulations include agreements on the evaluation and assessment of students. At the end of each academic year, the class council (teachers and the school principal) decides whether the student can progress. Government does not employ evaluation criteria or minimum cut-off grades. Unlike many other countries, there are no central examinations at the end of compulsory education (Rouw et al., 2016). Primary and secondary certificates are awarded by schools.

Quality triangle approach to education quality

While the public expect to hold government to account for education quality, it has had, historically, only weak governance over schools. The high degree of freedom of education makes steering by government difficult and school accountability mechanisms have not been strong (Högberg & Lindgren, 2021; Shewbridge et al., 2019). Government control of school quality has been limited to the work of the Inspectorate of Education (Janssens et al., 2016). Government will only recognise schools that participate in compulsory inspections. However, inspection cycles are long (up to ten years) and do not extend to school boards (OECD, 2023c). Inspection reports are public and include schools' strengths and weaknesses, whether they meet educational objectives and their quality assurance processes.

Much weight is placed on self-evaluation by schools. Accountability is viewed as an internal responsibility of schools, although there has been concern that their quality assurance systems are often underdeveloped (Flemish Inspectorate of Education, 2021; Shewbridge & Köster, 2021). Schools affiliate with umbrella organisations (Pedagogical Advisory Services) which offer curricula and assessment support. School collaboration is vertical, with umbrella organisations, rather than with other schools.

Collaboration between the three key bodies responsible for educational quality (schools, the Inspectorate and the Pedagogical Advisory Services) is known as the quality triangle approach (Shewbridge & Köster, 2021). All bodies follow the Reference Framework for Quality in Education (Flemish Ministry of Education and Training, 2018) to guide their work. The framework (known as the 'OK Framework') sets out good quality education via 37 indicators and seeks to respect the autonomy of schools (Flemish Inspectorate of Education, 2023). As such, the approach to supporting educational quality could represent a form of intelligent accountability as described by Sahlberg (2010).

Freedom of curriculum

The curriculum is defined by attainment targets – minimum objectives that students must meet at the end of a particular year or level of education: the goal being that all students receive a core curriculum and at least a minimal standard of educational quality (Simons et al., 2016). The targets have been the subject of extensive political and public debate (Loobuyck, 2021; Rouw et al., 2016; Simons et al., 2016). Some stakeholders have argued for limited targets, to give schools and teachers autonomy in how to operationalise them in their curriculum and pedagogy. The Inspectorate, however, has found the targets insufficiently concrete to support their judgements of educational quality (Simons et al., 2016).

The government set out to update the curriculum with the intention of introducing new competency-based attainment targets in 2014 (Flemish Parliament, 2014; OECD, 2021). New targets for the first stage of secondary education were implemented in 2019. Targets for the second and third stages of education were planned to be introduced in 2021 and 2023, respectively, but were challenged in court by the Catholic and Steiner education umbrella organisations. The Constitutional Court ruled that the new attainment targets violated the right to freedom of education. They were judged as being too detailed and extensive to allow for schools' own pedagogical approaches (Belgium Constitutional Court, 2022). Consequently, they were annulled in June 2022.

The Education Minister called the ruling 'a blow', arguing that as the quality of education had been falling for 20 years 'something must change and final attainment levels are the Government's only instrument for raising the bar' (Chini, 2022, para 7), illustrating the government's desire to strengthen central control of education. Nonetheless, the government began to co-develop with educational stakeholders less extensive and detailed minimum objectives (Flemish Parliament, 2023b). These have begun to be implemented but at the same time compulsory standardised tests are being introduced, offering an alternative mechanism of central control.

Standardised tests in primary and secondary education

In 2019, the government declared the need for standardised central assessments and went on to develop the first draft decree in 2021. This was contentious. Stakeholders feared that accountability mechanisms might limit school autonomy (Penninckx et al., 2017). They wished to prioritise and build trust (Shewbridge et al., 2019). The initiative was challenged by several stakeholders, and in particular by the Catholic umbrella organisation. The Constitutional Court balanced the risk that the tests might restrict the freedom of education against the supposed positive impacts of their introduction and judged that they were permissible. Despite fierce opposition (Penninckx, 2020), the decree for the tests was finally approved in April 2023 (Flemish Parliament, 2023a).

The government commissioned the ‘Flemish Research Center for Central Assessment in Education’ (the Steunpunt), a consortium of university researchers, to develop the tests. The tests, based on the attainment targets, are in Dutch (reading comprehension and writing) and mathematics. Tests in other subjects are planned for the future. The tests are compulsory for all students in Grades 4 and 6 in primary school, and 8 and 12 in secondary school. Grade 4 students are typically aged 9–10. Implementation will be gradual, with the first full administration for each grade happening between May 2024 and May 2027.

Policy rationale

While Flemish students have consistently performed above average on international tests, outcomes have declined over time. In the Trends in International Mathematics and Science Study (TIMSS), students’ average performance in mathematics and science has consistently decreased since 2003 (OECD, 2023c). The Progress in International Reading Literacy Study (PIRLS) showed a significant decrease in reading comprehension scores from 547 points in 2006 to 525 points in 2016 (Mullis et al., 2017, 2020). Since 2006, student performance in reading, mathematics and science has also been decreasing on the Programme for International Student Assessment (PISA; OECD, 2023b, 2023c; see Figure 1).

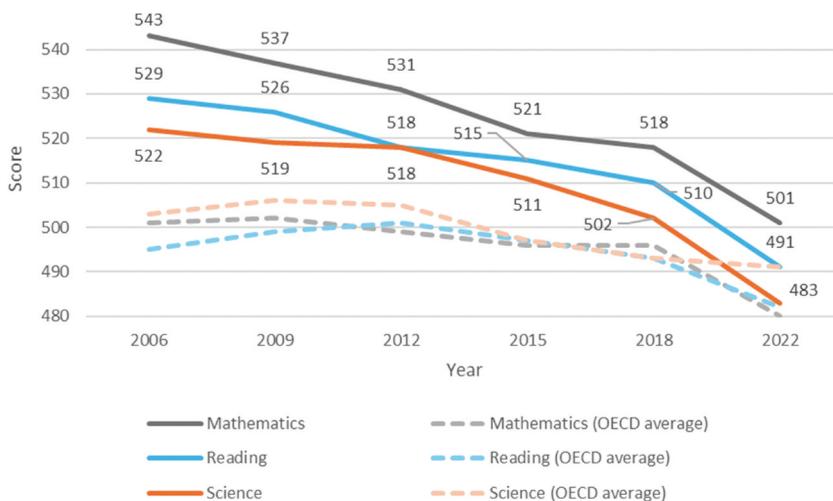


Figure 1. PISA results for Flanders compared to the OECD average, 2006 to 2022.

Although the average scores remain higher than for other regions in Belgium, the declines are larger than those observed across participating countries (OECD, 2023a).

The deterioration in performance is mirrored in the sample-based National Assessment Programme (the 'peilingen') at both primary and secondary levels. For example, even before the Covid-19 pandemic, a substantial proportion of students did not meet the mathematics attainment targets and outcomes had declined over time. Both international and national data also point to significant equity challenges, with students from lower socio-economic backgrounds performing significantly worse than their peers from more privileged backgrounds, with evidence of a larger attainment gap than in many other countries (OECD, 2023c).

The quality of education in Flanders may have diminished; after all, there has been a shortage of teachers (Pauwels et al., 2022). However, immigration has resulted in a more ethnically and culturally diverse population, which may also have contributed to declining outcomes. The percentage of students speaking a home language other than Dutch has risen (Janssens et al., 2016), as has the number of students with special educational needs (Flemish Parliament, 2019). Native speaker students have been shown to outperform language minority students on PISA (Agirdag & Vanlaar, 2018). There is, however, a dearth of research regarding the outcomes of students with special educational needs in PISA due to a lack of reliable public data on their participation (Gamazo et al., 2019).

Whatever the cause, declining international test results have '... led to a general feeling that there is something fundamentally wrong with the [education] system' (Penninckx, 2020, p. 40). While for a long time there was no sense of urgency to implement standardised testing as public trust in schools and the education system was high (Penninckx, 2020), a shared concern has built among stakeholders about educational quality (Shewbridge & Köster, 2021). Restoring the quality of education was one of the main debates in recent parliamentary elections (Chini, 2023b).

Persuasion is fundamental to politics and policy-making (Neustadt, 1960,1990) and an imperative for change was built around the test results (Backs et al., 2023; Denies et al., 2022). The website of the N-VA asserts 'Our quality of education has been under great pressure for years. For example, international comparative research – such as the PISA, TIMSS or PIRLS measurements – has shown declining performance for many years. Flemish Minister of Education Ben Weyts is taking up the fight against the declining quality of education with Flemish Tests' (New Flemish Alliance, 2023). Declining results have provided what Cochran-Smith and Fries (2001) called the 'evidentiary warrant' to construct the testing policy, which is characteristically presented as 'based solely on the empirical and certified facts of the matter and not embedded within or related to a particular agenda that is political or ideological' (p. 6).

The power of international tests to affect educational and assessment policy is well documented (see, for example, Baird et al., 2016; Lawn & Grek, 2012). Baird et al. (2016) showed that around the world, politicians have sometimes exaggerated the weakness of PISA performance to persuade the public of the need for reform. References to declining international test outcomes are politically powerful because they generate fears of falling behind in the global competition between knowledge economies and can be used to build coalitions among opposing interest groups (Steiner-Khamsi, 2014) and to gain discursive legitimacy (Fuller & Stevenson, 2019).

In the context of weakened economic growth, the tests are positioned as assessing knowledge and skills vital to the knowledge economy – a key part of the Flemish Reform Programme to increase economic prosperity (Government of Flanders, 2023). Less overt is the possibility that the tests might reinforce national culture in the context of immigration and a nationalistic agenda (cf. Broadfoot, 1999). Overall, the narrative is that the challenges of Flanders' social and economic context can be at least partly met by the panacea that is assessment.

Late to standardised testing

This case study invited reflection on whether Flanders had stood alone in its prolonged resistance to census-standardised tests. With 195 countries recognised by the United Nations and detailed information on many of their assessment systems difficult to obtain, establishing a definitive answer was challenging. Nonetheless, in international terms, Flanders was certainly a late adopter of standardised testing policy. Steiner-Khamsi (2014) argued that governments adopt policies that are internationally widespread only if it fits their domestic policy agenda. Besides, Kingdon (1995) observed that three ingredients are needed for a 'policy window' to open – recognition of a problem, an available solution and political willingness to create change (often generated when the political party in government changes). Declining international test results have highlighted the 'problem'. The political will to introduce testing came when the neoliberal N-VA became the dominant party in the government. All that was then needed to create a policy window was an available solution to the problem – standardised tests – which, in the words of the Minister for Education, would help to 'improve the quality of education in a targeted way' and to stop the education system 'sailing blind' (Weyts, 2023).

Nonetheless, this solution needed to be acceptable in the eyes of stakeholders. Acceptance of a policy is based on its perceived legitimacy, which is achieved through consensus building (Capano & Lippi, 2017). Belgium has a strong tradition of consensus building across a complex web of stakeholders (including four trade unions, a student association, three parent associations, three school networks, five umbrella organisations and five official bodies) slowing down the reform process. The policy of standardised testing was first mooted in 2014 and took a decade to get to the piloting phase. Acceptance was built slowly and carefully. This was likely worthwhile; as Gomendio and Wert (2023) noted, consensus over policy makes sustainability in the long term more likely.

The OECD played a significant role in building the legitimacy of the policy. It has been observed that, internationally, the OECD plays a 'soft governance' role (Niemann & Martens, 2018) and creates normative pressure – the expectation that nations will adapt to and learn from other systems (Martens & Jakobi, 2010). The OECD has widely employed its technical competencies to encourage reform agendas aimed to increase accountability (Grek, 2020). Verger, Fontdevila, et al. (2019) showed that the OECD promoted school autonomy with accountability solutions in many different settings in response to a broad range of problems, including issues with performance of the educational system and equity issues. Flanders has been the subject of several OECD reports that have noted the system's comparatively weak accountability mechanisms (see for example, Shewbridge et al., 2011, 2019, 2021). The Flemish Parliamentary (2019) policy briefing referred to the

work of the OECD many times and cited the OECD argument for constructive accountability in its rationale for standardised testing. Indeed, the Flemish government actively pursued opportunities for OECD input, perhaps recognising the legitimising role it plays.

Match to the Global Education Reform Movement (GERM) model

Policy research shows reform agendas should not be seen as top-down impositions, but as iterative constructions across various actors and pressures that have commonalities at the international level but which are more varied at national and local levels (see by way of illustration, Broadfoot's (2021) discussion of constants and contexts in pupils' experiences of education). The GERM is not a uniform model of education reform. The reshaping of education systems looks different between jurisdictions as it has many contextually specific drivers (Fuller & Stevenson, 2019). The GERM is organic, can evolve through different policy manifestations (Ball et al., 2017) and, as Sahlberg (2023) put it, education reforms often add 'their own peculiarities in implementation' (p. 3). Various agendas, priorities, and ideals underlie the development and enactment of educational assessment policies and practices (Ydesen, 2023). It is unlikely, then, that the education policy reforms in Flanders will perfectly fit the GERM prototype.

However, Ydesen and Dorn (2022) went further and argued that Sahlberg's GERM concept is too broad and skewed towards a portrayal of hegemonic authority. They claimed that governance mechanisms are subtle and cannot be explained as a one-way expansion of a global governance regime stemming from international organisations such as the OECD. They suggested that what appears to be global is, in fact, often a set of nationalistic ideologies that appear global at a distance. Indeed, Pollitt (1995) characterised New Public Management as a 'shopping basket' from which nations select policies as 'different countries have experienced different historical trajectories and seek to reform themselves within very different constitutional frameworks' (Pollitt, 1995, pp. 133–134).

Historically schools in Flanders have had high degrees of autonomy, and standardised testing has not been introduced in exchange for that autonomy. Rather than intended to stimulate school choice and market dynamics, the introduction of testing has been justified through quality concerns. In comparison to the GERM template, the introduction of the tests reflects a belated move by the government to create a different balance between autonomy and accountability. The policy does, however, leverage Broadfoot's (2021) social functions of assessment – the indirect control of the quality of education by quantification. The extent to which the tests might come to determine teaching and learning is discussed later. Next, we consider the design and stated purposes of the tests, and what these suggest about the implicit policy aims.

Test design

The tests are computer-based. The mathematics and reading comprehension tests will be adaptive – questions will vary in difficulty depending on students' previous responses in the test. Test items will not be released to schools following the test cycle. In reading, all students will be tested on the same competencies. The mathematics test will differ among students within and across schools. For example, the test taken by the secondary general

education stream will assess ten competencies related to solving mathematical problems such as number theory and algebra. Each year, each school will be randomly assigned three out of the ten competencies. Within a school, students will be randomly assessed on two of the three competencies assigned to their school (OECD, 2023c). Data will be shared with schools annually through a dashboard with results reported at the level of learning objectives and subject (Steunpunt, 2022).

Test purposes

The stated overarching objective of the tests is to strengthen and monitor the quality of education (Flemish Parliament, 2023a). The tests have aims at four levels: at the system level, to monitor the achievement of attainment targets and learning gains, providing an element of system-level quality assurance; at the school level, to inform external and internal quality assurance – schools and school boards will have access to their schools' contextualised feedback. Test results will also be used by the Pedagogical Advisory Services and by the Inspectorate of Education. At the class or student group level, the tests aim to allow reflection on the pedagogical didactic action of the school team. At the student level, the tests can contribute to the class council evaluation of students.

Stakeholders have highlighted the importance of contextualising test results for fair(er) comparisons and to aid interpretation (Molenberghs et al., 2023; OECD, 2023c). To this end a range of background indicators will be collected (e.g. mother's education level, home language, number of books at home). The Steunpunt will calculate learning gains at the system-, school- and student-level. Both raw and value-added results will be provided to schools (Flemish Parliament, 2023a), although explanations of precisely how the value-added measures will be calculated have not yet been provided and may prove contentious (as they have in England, for example [Leckie & Goldstein, 2017]).

Implicit purpose prioritisation

The tests have many stated purposes which have not been explicitly prioritised. This might be intended to maximise the use of the data, as well as to persuade sceptical interest groups of the policy benefits. The risk of this degree of purpose pluralism is that no purpose is achieved effectively (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; OECD, 2023c). There are features of the test design, however, that suggest an implicit prioritisation and that the use of test results by teachers has not been the central focus.

The sampling of mathematical content across schools will limit teachers' ability to evaluate the effectiveness of their teaching and the strengths and weaknesses of students. This is further complicated by the adaptive nature of the tests. As individual students are assessed on different competencies, using different items, teachers will be unable to make full comparisons across students. On the other hand, these design features can still ensure a holistic and comprehensive picture at the system level. The data will allow the Inspectorate to identify schools in need of support on the basis of repeated poor test performance or a decrease in learning gains.

Further, the confidentiality of the test items will impede teachers' abilities to use the data to inform their teaching. Teachers would benefit from having access to the test

questions to understand where students struggle and why – to support individual students, as well as to improve pedagogical practices or content coverage. This approach is followed in places such as Denmark and British Columbia (Canada), where assessments have a strong focus on pedagogical improvement (Government of British Columbia, n.d.; Wandall, 2017). Not sharing test items allows for test standards to be more easily maintained and for more accurate monitoring of trends over time, as changes to the content of the test are minimised. This again points towards a monitoring purpose. This choice, however, makes it less likely that tests will fully deliver on their promise of supporting teaching and learning, although it may impede unwanted teaching to the test. The design of the tests risks the production of what Broadfoot (1996) termed ‘dead data’ - information unlikely to be of any great use in informing teaching.

While the tests are not optimally designed to support improvements in educational quality, they may nonetheless have a strong effect on what happens in the classroom. This will depend on the perceived stakes of the tests.

Intended low stakes

Stakeholders were very concerned that the test results might be used to create school rankings (league tables). They have been acutely aware of the unintended consequences of testing in other countries such as England (Penninckx et al., 2017). Penninckx (2020) made the case for ‘low-stakes’ standardised testing in Flanders, in which results are only communicated to schools and not used in decisions to award certificates to students. Indeed, the stated policy was that the tests will be low stakes so as not to have a negative washback on teaching and learning (OECD, 2023c). The Flemish government has legislated to ensure that the results of a school are not made public and to prevent school rankings from being created (Weyts, 2023). The government has stated that ‘... the tests should be supportive for schools and not a means for school accountability. Since the tests are low-stakes for schools, the schools will not engage in strategic behaviour to artificially increase their scores’ (Flemish Government, 2023).

In the longer term, the government’s determination to prevent school rankings may flounder in the face of pressure from stakeholders including parents, as was the case in Israel, for example (Feniger et al., 2016). Regardless, experience shows that it is impossible to control the stakes associated with any assessment. Negative effects can emerge and grow, as demonstrated by national assessments in primary schools in England (Perryman et al., 2011; Wyse & Torrance, 2009), the National Student Achievement Tests in Lithuania (Raudienė et al., 2022) and Israel’s Meitzav tests (Feniger et al., 2016). Small changes in the use of assessment data or how assessment results are communicated can impact the perceived stakes (Feniger et al., 2016). After all, the power of tests to affect behaviour is a perceptual phenomenon: ‘If students, teachers, or administrators believe that the results of an examination are important, it matters very little whether this is really true or false – the effect is produced by what individuals perceive to be the case’ (Madaus, 1988, p. 35).

Moreover, in common with many European education systems (Ehren & Shackleton, 2016; Timmermans et al., 2015), school-level results will be shared with the Inspectorate of Education as part of a risk-based inspection regime. This is somewhat at odds with the government’s claim that the tests will not be used for accountability purposes. The Inspectorate will use the test results in combination with other information (such as the

student intake characteristics of the school) to target visits (Flemish Government, 2023; OECD, 2023c). When this happens, teachers and schools are unlikely to view test outcomes as low stakes. Too strong and narrow a focus on test scores in risk-based inspections can lead to gaming and manipulation of scores (Ehren & Honingh, 2011). Indeed, despite the previous inspection context in Flanders being ‘low stakes’, schools engaged in strategic activities to produce a positive school image, and staff experienced negative emotional effects (Penninckx et al., 2016). While stakeholders have expressed concerns that test data may come to overly influence the Inspectorate’s evaluations, the Inspectorate have emphasised their continued use of the OK Framework which embodies a wide understanding of educational quality (OECD, 2023c). Indeed, Shewbridge and Köster (2021) argued that the test data should be placed within the strengths of the current accountability system, which focuses on dialogue and understanding the links between outcomes and improving practice.

Likely short- and long-term effects of standardised testing

The overarching goal of the standardised testing policy is to improve the quality of education. Next, we draw on international research evidence to reflect on whether standardised testing will have the envisioned positive effects. There is debate about whether similar reforms around the world have improved education (for example, see Sahlberg [2023] for a pessimistic account, and Gomendio and Wert [2023] for a more positive interpretation of the evidence). Other scholars question whether it is ever possible to evaluate the success or failure of any such reform (e.g. Corsi, 2023).

The need to support teachers’ data use to improve education quality

Setting aside the suboptimal design of the tests for use by teachers, for the introduction of the tests to improve educational quality teachers will need to be able to interpret the data and turn them into actionable plans. Historically, the use of performance data by schools has been weak in Flanders – used mainly in an ad hoc rather than strategic manner (Shewbridge et al., 2011). For example, in a study of primary school teachers’ decision-making processes regarding grade retention, Vanlommel et al. (2017) found that decisions were based mostly on intuitive expertise and rarely on data.

Given this, teachers want to deepen their professional skills in the use of test data (Shewbridge & Köster, 2021). Steunpunt (2022) has committed to the development of online courses on information literacy (how to interpret results) and educational development (how to improve educational quality using results). It is unlikely that these courses alone, even if enthusiastically taken up, will ensure that the data is used to good effect. A much broader package of support and training is likely to be needed. For example, research highlights the importance of the role of school leaders in creating a culture of positive data use (Schildkamp et al., 2019) and the importance of teacher collaboration, as solutions to problems and consequent actions based on analysis of data are rarely self-evident (Mandinach & Schildkamp, 2021). Further, the use of test data in conjunction with other diverse data sources would be best and would inform teachers about the whole student, but again this would require training and support (Mandinach & Gummer, 2016). In summary, without a range of training and support for all users of

standardised test outcomes, their introduction is unlikely to be effective in improving educational quality. Such training and support will need to be designed in the knowledge of the potential for unintended consequences of testing.

Impact on teaching, learning and educational culture

The government has stated that the tests are intended to be low stakes and has perhaps sought to manage the risk of teaching to the test by not releasing items. The tests were also designed to require students to apply insight and reasoning skills in the belief that instrumental teaching to the test would not be rewarded with higher test scores (The Brussels Times, 2022). However, while some tests may be easier to train students for than others, research shows that even aptitude and intelligence test scores can be improved with coaching (Freund & Holling, 2011; Kulik et al., 1984).

Given that the stakes are unlikely to be perceived as unequivocally low, some form of teaching to the test is highly likely. Arguably, the impact of the tests on teaching and learning may be greater than otherwise would be the case because of the sparsity of the centrally defined curriculum. Even without the release of test items, teachers will, over time, become familiar with the format and content of the tests and modify their teaching in response. Indeed, there are anecdotal reports of teachers photographing test items during the piloting phase. There is a risk that over time, teachers will devote greater time to the tested areas of the curriculum, omit untested topics from teaching, and even go so far as to coach students on the style of test items and test-taking heuristics (Koretz, 2005; Pollard et al., 2001; Russell et al., 2009).

Teachers and school leaders may change the way they think about their profession and about education (Ball, 2003; Taylor, 2023). The representation of quality and improvement through test data risks the conflation of quality and its representation. Efforts may become concentrated on what can be measured and acted on, rather than on what is truly valued (Lawn & Grek, 2012). Teachers' understanding of their students could become dominated by test scores (Booher-Jennings, 2005). There is a risk that standardised testing is the precursor to a new performative education culture (Ball, 2003). Comparisons between schools may become the lens through which teachers and school leaders look at their own school and make decisions. What seems certain is that, over time, the quantification of education newly created by the test data will open up Flanders' education system to further policy reforms of various kinds, as comparisons raise questions and demand action (Mangez & Broeck, 2021). As Corsi (2023) put it, 'Reforms generate reforms' (p. 371).

Intelligent accountability

Extreme negative consequences of standardised testing are not, however, a foregone conclusion. As described earlier, Flanders' approach to supporting education quality, the OK Framework, could be characterised as a form of intelligent accountability. Sahlberg (2010) argued that intelligent accountability can include standardised testing. Intelligent approaches combine internal accountability (e.g. through school self-evaluation) with levels of external accountability, including monitoring via assessment, although the efficiencies of sample, rather than census, testing are emphasised.

For such an approach to accountability to succeed in Flanders, test data must not come to dominate the shared vision of educational quality embodied by the OK Framework. The framework is strongly embedded in the education system and represents a wide and largely shared vision of educational quality. The framework has the development of the learner at the centre and emphasises the needs of *all* learners. It includes indicators related to pupil wellbeing, innovative practice, professional dialogue and reflection. Only five of 37 indicators of quality in the framework are directly or indirectly associated with assessment data (OECD, 2023c).

As envisioned by Shewbridge and Köster (2021), rich, nuanced sources of evidence of quality need to continue to be valued, and test data need to be used to stimulate critical self-reflection and professional dialogue. The levels of autonomy and trust historically granted to schools indicate a tradition of trust. They have given rise to a strong sense of teacher professionalism and teachers feel a sense of ownership of their classrooms (Penninckx, 2020). This bodes well for teachers continuing to view accountability as their responsibility, internal to schools, rather than a top-down imposition.

Dependability of test data over time

Whether or not the introduction of standardised tests has the intended impact on educational quality, the test data itself may give a positive picture. This is because as the stakes associated with assessment systems grow, even well-designed tests can lose dependability (Koretz, 2008). Over time, the changing behaviours of students, teachers and school leaders reduce test validity and therefore the dependability of the scores (Stobart, 2008). Test scores increase year on year but the changes do not reflect genuine improvements in students' attainment. The improvements are specific to the test and fail to generalise to performance in related tests or other contexts (Anastasi, 1981; Cuff et al., 2019; Koretz, 2005). There is a risk that, particularly in the early years of testing when the opportunity to increase test familiarity is at its greatest, there will be improvements in test scores which may be misinterpreted as indicating improved educational quality and as a policy success.

Conclusion

In light of falls in international test results, the Flemish government has embarked on the introduction of compulsory standardised testing. The policy changes in Flanders vary from the archetypal GERM template. Standardised testing has not been introduced in exchange for greater autonomy and is not intended to stimulate school choice and market dynamics. However, a fundamental feature of the GERM – the use of assessment to govern education – is evident. The government wishes to create a different balance between school autonomy and accountability.

An attempt to gain more control over what is taught in schools by setting a more detailed centrally controlled curriculum failed on the basis that it unduly constrained the freedom of education. However, as Broadfoot (2021) established, the introduction of standardised tests is an alternative, insidious and powerful source of control. It is telling that the design of the tests has likely prioritised system and school-level monitoring over teacher use of the data for formative purposes. The policy of

standardised testing gives the appearance of freedom but, in reality, there will be an increase in central control gained in a manner which may prove hard to challenge legally.

The organic nature of the effects of policy make outcomes hard to predict, with history, culture, values, customs and structures all affecting the likely success of reforms. Features of Flemish educational culture, in particular the powerful belief in freedom of education, may insulate the system from some of the most extreme perverse consequences of compulsory standardised testing documented elsewhere. Over time, however, Flanders' education system will almost certainly face more change as the use of standardised test data raises questions which demand action.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Michelle Meadows is Associate Professor in Educational Assessment at Oxford University. Her research interests span most elements of high-stakes assessments and qualifications. She has published research into marking, standard setting, comparability, qualification structures and malpractice. Her previous roles include Deputy Chief Regulator at Ofqual, England's exams regulator and Director of AQA's Centre for Education Research and Practice.

Inés Sanguino is an Advisor in the Home Affairs, Social Cohesion and Education team at the Behavioural Insights Team in London. She previously worked with the Implementing Education Policies Team at the OECD. She has also worked with organisations such as What Works for Children in Social Care and Unlocked Graduates. She has collaborated on various projects at the Junior Researcher, King's College London and the University of Oxford. Inés completed a BSc in Psychology, and an MPhil in Evidence-Based Social Intervention and Policy Evaluation as a 'La Caixa' Scholar.

ORCID

Michelle Meadows  <http://orcid.org/0009-0007-5200-6331>

References

- Adam, I., & Xhardez, C. (2024). Chapter, 9: How sub-state nationalism and immigrant integration policies entwine over time: A spotlight on Flanders (Belgium). In M. Samers & J. Rydgren (Eds.), *Migration and nationalism* (pp. 174–196). Edward Elgar Publishing.
- Agirdag, O., & Vanlaar, G. (2018). Does more exposure to the language of instruction lead to higher academic achievement? A cross-national examination. *International Journal of Bilingualism*, 22(1), 123–137. <https://doi.org/10.1177/1367006916658711>
- Alderson, C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115–129. <https://doi.org/10.1093/applin/14.2.115>
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Anastasi, A. (1981). Coaching, test sophistication, and developed abilities. *The American Psychologist*, 36(10), 1086–1093. <https://doi.org/10.1037/0003-066X.36.10.1086>

- Andeweg, R. B. (2019). Consociationalism in the low countries: Comparing the Dutch and Belgian experience. *Swiss Political Science Review*, 25(4), 1–18. <https://doi.org/10.1111/spsr.12361>
- Backs, J., de Fraine, B., Sterck, M., D'haese, J., & Ballet, K. (2023). Evidence use in implementing standardised testing in the Flemish Community of Belgium. In *Who really cares about using education research in policy and practice? Developing a culture of research engagement* (pp. 95–106). OECD Publishing.
- Baird, J.-A. J., Hopfenbeck, S., Isaacs, T. N., Stobart, T., Sprague, T., Stobart, G., & Yu, G. (2016). On the supranational spell of PISA in policy. *Educational Research*, 58(2), 121–138. <https://doi.org/10.1080/00131881.2016.1165410>
- Ball, S. J. (1994). *Education reform: A critical and post-structural approach*. Open University Press.
- Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, 18(2), 215–228. <https://doi.org/10.1080/0268093022000043065>
- Ball, S. J., Junemann, C., & Santori, D. (2017). *Edu.net. Globalisation and education policy mobility*. Routledge.
- Belga News Agency. (2023, August 30). *Back to school: Dutch only in classrooms and playgrounds, says minister*. <https://www.belganewsagency.eu/back-to-school-dutch-only-in-classrooms-and-playgrounds-says-flemish-minister>
- Belgium Constitutional Court. (2022, June 16). *Press release judgement 82/2022*. <https://www.const-court.be/public/n/2022/2022-082n-info.pdf>
- 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>
- Broadfoot, P. (1999). The role of educational assessment in England and France. *The Tocqueville Review*, 20(2), 9–24. <https://doi.org/10.3138/tr.20.2.9>
- Broadfoot, P. (2021). *The sociology of assessment: Comparative and policy perspectives: The selected works of Patricia Broadfoot*. Routledge.
- Broadfoot, P. M. (1996). Do we really need to write it all down? Managing the challenge of national assessment at KS1 and KS2. In P. Croll (Ed.), *Teachers, pupils and primary schooling* (pp. 69–85). Cassell.
- The Brussels Times*. (2022, July 16). Flanders approves plans for centralised exams in primary and secondary schools. <https://www.brusselstimes.com/256409/flanders-approves-plans-for-centralised-exams-in-primary-and-secondary-schools>
- Capano, G., & Lippi, A. (2017). How policy instruments are chosen: Patterns of decision makers' choices. *Policy Sciences*, 50(2), 269–293. <https://doi.org/10.1007/s11077-016-9267-8>
- Chini, M. (2022, June 17). *Constitutional court annuls Flanders' attainment targets for secondary school*. *The Brussels times*. <https://www.brusselstimes.com/239947/constitutional-court-annuls-flanders-entry-targets-for-secondary-school>
- Chini, M. (2023a, August 30). *'Makes all other subjects possible': Flanders to enforce grades for Dutch in kindergarten*. *The Brussels times*. <https://www.brusselstimes.com/668113/makes-all-other-subjects-possible-flanders-to-enforce-grades-for-dutch-in-kindergarten>
- Chini, M. (2023b, December 9). *Belgian education under fire after 'bombshell' school performance levels*. *The Brussels Times*. <https://www.brusselstimes.com/829248/belgian-education-under-fire-after-bombshell-school-performance-levels>
- Cochran-Smith, M., & Fries, M. K. (2001). Sticks, stones and ideology: The discourse of reform in teacher education. *Educational Researcher*, 30(8), 3–15. <https://doi.org/10.3102/0013189X030008003>
- Corsi, G. (2023). Education reform as a global phenomenon. In P. Mattei, X. Dumay, E. Mangez, & J. Behrend (Eds.), *The Oxford handbook of education and globalization* (pp. 367–384). Oxford University Press.
- Cuff, B. M., Meadows, M., & Black, B. (2019). An investigation into the Sawtooth Effect in secondary school assessments in England. *Assessment in Education Principles, Policy & Practice*, 26(3), 321–339. <https://doi.org/10.1080/0969594X.2018.1513907>

- De Coninck, D., & Joris, W. (2021). What a difference a year makes: Changes in refugee threat perceptions in Flanders, Belgium. *Belgium Sage Open*, 11(3). <https://doi.org/10.1177/21582440211036884>
- De Groof, J., van de Donk, W., Lauwers, G., de Goede, P., & Verhappen, T. (2010). Reflections on religion & education in the Netherlands and Flanders. In *Religion, beliefs, philosophical convictions and education conference* (pp. 7–9). European Commission.
- Denies, K., Vanbuel, M., Aesaert, K., Van Keer, H., Adriaens, V., Erauw, I., & von Davier, M. (2022). *Belgium (Flemish)*. Pirls 2021 Encyclopedia. <https://pirls2021.org/belgium-flemish/>
- Ehren, M. C. M., & Honingh, M. (2011). Risk-based school inspections in the Netherlands: A critical reflection on intended effects and causal mechanisms. *Studies in Educational Evaluation*, 37(4), 239–248. <https://doi.org/10.1016/j.stueduc.2012.02.001>
- Ehren, M. C., & Shackleton, N. (2016). Risk-based school inspections: Impact of targeted inspection approaches on Dutch secondary schools. *Educational Assessment, Evaluation and Accountability*, 28(4), 299–321. <https://doi.org/10.1007/s11092-016-9242-0>
- Feniger, Y., Israeli, M., & Yehuda, S. (2016). The power of numbers: The adoption and consequences of national low-stakes standardised tests in Israel. *Globalisation, Societies & Education*, 14(2), 183–202. <https://doi.org/10.1080/14767724.2015.1010438>
- Flemish Government. (2023). *Flemish tests*. Retrieved November 9, 2023, from <https://www.vlaanderen.be/en/studying/flemish-tests>
- Flemish Inspectorate of Education. (2021). *Education mirror 2021*. Retrieved November 19, 2023, from <https://publicaties.vlaanderen.be/view-file/41674>
- Flemish Inspectorate of Education. (2023). *OK reference framework*. <https://english.onderwijsinspectie.nl/>
- Flemish Ministry of Education and Training. (2018). *OK education quality, Flemish ministry of education and training*. Retrieved November 15, 2023, from <https://www.onderwijsinspectie.be/nl/het-referentiekader-voor-onderwijskwaliteit-het-ok>
- Flemish Parliament. (2014). *Education policy note 2014–2019*. Retrieved November 16, 2023, from <https://www.vlaamsparlement.be/nl/parlementaire-documenten/parlementaire-initiatieven/934626>
- Flemish Parliament. (2019). *Policy brief: Education and training policy paper*. Retrieved November 16, 2023, from <https://docs.vlaamsparlement.be/pfile?id=1498215>
- Flemish Parliament. (2023a). *Decree on the Flemish tests in education*. Retrieved November 14, 2023, from <https://beslissingenvlaamseregering.vlaanderen.be/document-view/644A3479CA1CB15B58CF5300>
- Flemish Parliament. (2023b). *Proposal of decree. About the educational goals for the second and third stages of secondary education*. Retrieved November 24, 2023, from <https://docs.vlaamsparlement.be/pfile?id=1970520>
- Freund, P. A., & Holling, H. (2011). How to get really smart: Modeling retest and training effects in ability testing using computer-generated figural matrix items. *Intelligence*, 39(4), 233–243. <https://doi.org/10.1016/j.intell.2011.02.009>
- Fuller, K., & Stevenson, H. (2019). Global education reform: Understanding the movement. *Educational Review*, 71(1), 1–4. <https://doi.org/10.1080/00131911.2019.1532718>
- Gaasendam, C. R., Abts, K., Swyngedouw, M., & Meuleman, B. (2021). Lost connection? The attitudinal and ideological (in) congruence of social democracy's elites, members and voters in Flanders-Belgium. *Acta Politica*, 56(3), 395–415. <https://doi.org/10.1057/s41269-020-00156-6>
- Gamazo, A., Martínez-Abad, F., Olmos-Migueláñez, S., & Rodríguez-Conde, M. J. (2019, October). Assessment of students with special education needs in PISA: Participation guidelines and data treatment. *Proceedings of the seventh international conference on technological ecosystems for enhancing multiculturalism* (pp. 187–191).
- Godfrey-Faussett, T., & Baird, J. (2025). What does success mean to you? Negotiating individual definitions of educational success within an examination dominated regime of truth. *Oxford Review of Education*, 51(2), pp. 178–201.
- Gomendio, M., & Wert, J. I. (2023). *Dire straits-education reforms: Ideology, vested interests and evidence*. Open Book Publishers.

- Government of British Columbia. (n.d.). *Foundation Skills Assessment (FSA)*. Retrieved November 15, 2023 from <https://www2.gov.bc.ca/gov/content/education-training/k-12/administration/program-management/assessment/foundation-skills-assessment#brochures>
- Government of Flanders. (2023). *Flemish reform Programmer 2023*. <https://financeflanders.be/wp-content/uploads/2023/05/Flemish-Reform-Programme-2023.pdf>
- Grek, S. (2020). Facing “a tipping point”? The role of the OECD as a boundary organisation in governing education in Sweden. *Education Inquiry*, 11(3), 175–195. <https://doi.org/10.1080/20004508.2019.1701838>
- Hamilton, L. S., Stecher, B. M., & Klein, S. P. (2002). *Making sense of test-based accountability in education*. RAND Corporation.
- 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>
- Janssens, D. D., Van Damme, P. J., & Bellens, K. (2016). *PIRLS, 2016 encyclopaedia: Belgium, Flemish community*. [https://pirls2016.org/wp-content/uploads/encyclopedia-pirls/downloadcenter/3.%20Country%20Chapters/Belgium%20\(Flemish\).pdf](https://pirls2016.org/wp-content/uploads/encyclopedia-pirls/downloadcenter/3.%20Country%20Chapters/Belgium%20(Flemish).pdf)
- Kingdon, J. (1995). *Agendas, alternatives, and public policies* (2nd ed.). Longman.
- Koretz, D. (2005). Alignment, high stakes, and the inflation of test scores. *Teachers College Record*, 107(14), 99–118. <https://doi.org/10.1177/016146810510701405>
- Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Harvard University Press.
- Kulik, J. A., Bangert-Drowns, R. L., & Kulik, C. C. (1984). Effectiveness of coaching for aptitude tests. *Psychological Bulletin*, 95(2), 179–188. <https://doi.org/10.1037/0033-2909.95.2.179>
- Lawn, M., & Grek, S. (2012). *Europeanizing education: Governing a new policy space*. Symposium Books.
- Leckie, G., & Goldstein, H. (2017). The evolution of school league tables in England 1992–2016: ‘contextual value-added’, ‘expected progress’ and ‘progress 8’. *British Educational Research Journal*, 43(2), 193–212. <https://doi.org/10.1002/berj.3264>
- Lijphart, A. (1977). *Democracy in plural societies: A comparative exploration*. Yale University Press.
- Loobuyck, P. (2021). The policy shift towards citizenship education in Flanders. How can it be explained? *Journal of Curriculum Studies*, 53(1), 65–82. <https://doi.org/10.1080/00220272.2020.1820081>
- Madaus, G. F. (1988). The distortion of teaching and testing: High-stakes testing and instruction. *Peabody Journal of Education*, 65(3), 29–46. <https://doi.org/10.1080/01619568809538611>
- Maddens, B., Muyters, G., Wolfs, W., & Van Hecke, S. (2020). The European Union in the narratives of secessionist parties: Lessons from Catalonia, Flanders, and Scotland. In R. D. Griffiths & D. Muro (Eds.), *Strategies of secession and counter-secession* (pp. 122–142). Rowen & Littlefield in partnership with ECPR Press.
- Maly, I. (2016). ‘Scientific’ nationalism: N-VA and the discursive battle for the Flemish nation. *Nations and Nationalism*, 22(2), 266–286. <https://doi.org/10.1111/nana.12144>
- Mandinach, E. B., & Gummer, E. S. (2016). *Data literacy for educators: Making it count in teacher preparation and practice*. Teachers College Press.
- Mandinach, E. B., & Schildkamp, K. (2021). Misconceptions about data-based decision making in education: An exploration of the literature. *Studies in Educational Evaluation*, 69, 100842. <https://doi.org/10.1016/j.stueduc.2020.100842>
- Mangez, E., & Broeck, P. V. (2021). The construction of worlds: On the function and fiction of quantified data in education policy. In S. Grek, C. Maroy, & A. Verger (Eds.), *World yearbook of education 2021* (1st ed., Vol. 1, pp. 25–42). Routledge.
- Martens, K., & Jakobi, A. (2010). *Mechanisms of OECD governance: International incentives for national policy-making?* Oxford University Press.
- Molenberghs, G., van Gasse, R., de Maeyer, S., & Vanhoof, J. (2023). *Feedback use of Flemish tests. Survey research among school leaders and teachers*. Steunpunt for the Flemish tests. <https://steunpunttoetsen.be/wp-content/uploads/2023/04/Feedbackgebruik-uit-centrale-toetsen.-Survey-onderzoek-bij-schoolleiders-en-leraren.pdf>

- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS, 2016 international results in reading*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement. <https://files.eric.ed.gov/fulltext/ED580353.pdf>
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS, 2019 international results in mathematics and science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center. <https://timssandpirls.bc.edu/timss2019/international-results/>
- Neustadt, R. E. (1960). *Presidential power*. Wiley.
- Neustadt, R. E. (1990). *Presidential power and the modern presidents. The politics of leadership from Roosevelt to Reagan*. The Free Press.
- New Flemish Alliance. (2023). *Flemish tests approved: "Small revolution in our schools"*. <https://www.benweyts.be/nieuws/vlaamse-toetsen-goedgekeurd-kleine-revolutie-in-onze-scholen>
- Niemann, D., & Martens, K. (2018). Soft governance by hard fact? The OECD as a knowledge broker in education policy. *Global Social Policy, 18*(3), 267–283. <https://doi.org/10.1177/1468018118794076>
- OECD. (2018). *Education at a glance, 2018: OECD indicators*. OECD Publishing.
- OECD. (2021). *Teachers' professional learning study: Diagnostic report for the Flemish community of Belgium*. OECD Education Policy Perspectives, 31, OECD Publishing.
- OECD. (2023a). *Decline in educational performance only partly attributable to the COVID-19 pandemic*. <https://www.oecd.org/newsroom/decline-in-educational-performance-only-partly-attributable-to-the-covid-19-pandemic.htm>
- OECD. (2023b). *PISA, 2022 results (volume I): The state of learning and equity in education*. OECD Publishing.
- OECD. (2023c). *Strengthening the design and implementation of the standardised student assessment reform of the Flemish community of Belgium*. OECD Education Policy Perspectives, 72, OECD Publishing.
- Olmedo, A., & Wilkins, A. (2017). Governing through parents: A genealogical enquiry of education policy and the construction of neoliberal subjectivities in England. *Discourse: Studies in the Cultural Politics of Education, 38*(4), 573–589. <https://doi.org/10.1080/01596306.2015.1130026>
- Pauwels, K., Van Loon, H., Tanghe, E., & Schelfhout, W. (2022). Motives of student teachers in academic teacher education for secondary education: Research in Flanders (Belgium) on the motivation to become and to remain a teacher. *Education Sciences, 12*(10), 724. <https://doi.org/10.3390/educsci12100724>
- Penninckx, M. (2020). Quality education without central examinations: The case of Flanders. *Management in Education, 34*(1), 40–43. <https://doi.org/10.1177/0892020619881045>
- Penninckx, M., Maeyer, J., Vanhoof, D., & Van Petegem, P. (2017). Delphi study on standardized systems to monitor student learning outcomes in Flanders: Mechanisms for building trust and/or control? *Studia paedagogica, 22*(2), 9–31. <https://doi.org/10.5817/SP2017-2-2>
- 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>
- Pollard, A., Broadfoot, P., McNess, E., Osborn, M., & Triggs, P. (2001). *What teachers do*. Bloomsbury Publishing Plc.
- Pollitt, C. (1995). Justification by works or by faith? Evaluating the new public management. *Evaluation, 1*(2), 133–154. <https://doi.org/10.1177/135638909500100202>
- Raudienė, I., Kaminskienė, L., & Cheng, L. (2022). The education and assessment system in Lithuania. *Assessment in Education Principles, Policy & Practice, 29*(3), 383–394. <https://doi.org/10.1080/0969594X.2022.2103516>
- Rouw, R., Fuster, M., Burnsi, T., & Brandt, M. (2016). *United in diversity: A complexity perspective on the role of attainment targets in quality assurance in Flanders*. OECD Education Working Papers, 139. OECD Publishing.

- Russell, M., Madaus, G., & Higgins, J. (2009). *The paradoxes of high stakes testing: How they affect students, their parents, teachers, principals, schools, and society*. Information Age Publishing, Inc.
- Sahlberg, P. (2010). Rethinking accountability in a knowledge society. *Journal of Educational Change*, 11(1), 45–61. <https://doi.org/10.1007/s10833-008-9098-2>
- Sahlberg, P. (2016). The global educational reform movement and its impact on schooling. In K. Mundy, A. Green, B. Lingard, & A. Verger (Eds.), *The handbook of global education policy* (pp. 128–144). Wiley-Blackwell.
- Sahlberg, P. (2023). Trends in global education reform since the 1990s: Looking for the right way. *International Journal of Educational Development*, 98, 102748. <https://doi.org/10.1016/j.ijedudev.2023.102748>
- Schildkamp, K., Poortman, C. L., Ebbeler, J., & Pieters, J. M. (2019). How school leaders can build effective data teams: Five building blocks for a new wave of data-informed decision making. *Journal of Educational Change*, 20(3), 283–325. <https://doi.org/10.1007/s10833-019-09345-3>
- Shewbridge, C., Fuster, M., & Rouw, R. (2019). *Constructive accountability, transparency and trust between government and highly autonomous schools in Flanders*. OECD Publishing.
- Shewbridge, C., Hulshof, M., Nusche, D., & Stoll, L. (2011). *School evaluation in the Flemish community of Belgium 2011*. OECD Reviews of Evaluation and Assessment in Education, OECD Publishing.
- Shewbridge, C., & Köster, F. (2021). *Educational research and innovation promoting education decision makers' use of evidence in Flanders*. OECD Publishing.
- Simons, M., Kelchtermans, G., Leysen, J., & Vandenbroeck, M. (2016). *Attainment targets in action. Mapping the operation, effects and future use of a Flemish educational policy instrument*. Laboratory for Education and Society & Centre for Educational Policy and Innovation & Teacher Training. <https://publicaties.vlaanderen.be/view-file/20165>
- Steiner-Khamsi, G. (2014). Cross-national policy borrowing: Understanding reception and translation. *Asia Pacific Journal of Education*, 34(2), 153–167. <https://doi.org/10.1080/02188791.2013.875649>
- Steunpunt. (2022). *Tests for educational development. Key principles for the central assessments in Flanders*. Support Center for Central Testing in Education. <https://steunpunttoetsen.be/>
- Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Routledge.
- Taylor, L. A. (2023). "If I was better at managing all this": The role of neoliberal logic in framing one teacher's narratives about accountability. *Teaching & Teacher Education*, 121, 103944. <https://doi.org/10.1016/j.tate.2022.103944>
- Timmermans, A. C., De Wolf, I. F., Bosker, R. J., & Doolaard, S. (2015). Risk-based educational accountability in Dutch primary education. *Educational Assessment, Evaluation and Accountability*, 27(4), 323–346. <https://doi.org/10.1007/s11092-015-9212-y>
- Tourne, J., Roose, R., Devlieghere, J., & Bradt, L. (2024). From crisis to commons? Exploring the potential of the commons via two secondary education case studies in Flanders. In G. Cappello, M. Siino, N. Fernandes, & M. Arciniega-Cáceres (Eds.), *Educational commons: Democratic values, social justice and inclusion in education* (pp. 13–27). Springer.
- Vanlommel, K., Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2017). Teachers' decision-making: Data based or intuition driven? *International Journal of Educational Research*, 83, 75–83. <https://doi.org/10.1016/j.ijer.2017.02.013>
- Verger, A., Fontdevila, C., & Parcerisa, L. (2019). Constructing school autonomy with accountability as a global policy model: A focus on OECD governance mechanisms. In C. Ydesen (Ed.), *The OECD's historical rise in education: The formation of a global governing complex* (pp. 219–243). Palgrave Macmillan.
- Verger, A., Parcerisa, L., & Fontdevila, C. (2019). The growth and spread of large-scale assessments and test-based accountabilities: A political sociology of global education reforms. *Educational Review*, 71(1), 5–30. <https://doi.org/10.1080/00131911.2019.1522045>
- Wandall, J. (2017). National tests in Denmark: CAT as a pedagogic tool. *The Forum*, 59(2), 145. <https://doi.org/10.15730/forum.2017.59.2.145>
- Weyts, B. (2023). *Green light for Flemish tests: Test with 21, 000 students*. NVA. Retrieved 9 November 2023, from <https://www.benweyts.be/nieuws/groen-licht-voor-vlaamse-toetsen-test-bij-21000-leerlingen>

- Wyse, D., & Torrance, H. (2009). The development and consequences of national curriculum assessment for primary education in England. *Educational Research*, 51(2), 213–228. <https://doi.org/10.1080/00131880902891479>
- Ydesen, C. (2023). Educational assessment – trends in research, policy and practice. In R. J. Tierney, F. Rizvi, & K. Ercikan (Eds.), *International encyclopedia of education* (4th ed., pp. 103–107). Elsevier Ltd. <https://doi.org/10.1016/B978-0-12-818630-5.09033-3>
- Ydesen, C., & Dorn, S. (2022). The No Child left behind Act in the global architecture of educational accountability. *History of Education Quarterly*, 62(3), 268–290. <https://doi.org/10.1017/heq.2022.11>
- Yin, R. K. (2009). *Case study research: Design and methods* (Vol. 5). Sage Publications.