Strengthening the design and implementation of the standardised student assessment reform of the Flemish Community of Belgium

The Flemish Community of Belgium has one of the OECD's most devolved education systems, with schools enjoying a high degree of autonomy and parents benefitting from free school choice. But there are concerns among Flemish education stakeholders about the overall quality of education. For example, while 15-year-old Flemish students have repeatedly scored higher in reading, mathematics and science than the OECD average in PISA (the OECD's Programme for International Student Assessment), student performance has been decreasing, with more students failing to reach basic proficiency levels in these subjects than before.

In recent years, the Flemish government has initiated a series of reforms to help strengthen the overall quality of education. These include the introduction of full cohort standardised student assessments in Dutch language and mathematics for primary and secondary education. Students will take the assessments in Grades 4 and 6 at the primary level and in Grades 8 and 12 at the secondary level starting in 2024. This ambitious and innovative reform aims to strengthen and monitor the quality of education, with student results envisaged to inform improvement efforts at different levels of the system.

As a follow up to the support provided by the OECD Strategic Education Governance Project, the Flemish Department of Education and Training (DoET) asked the OECD Implementing Education Policies Project team to support the successful design and implementation of the standardised student assessment reform of the Flemish Community of Belgium. This report presents the analysis, key findings and recommendations to help realise this objective.

The report was prepared by Michelle Meadows, consultant at the University of Oxford; Inge de Wolf, consultant at Maastricht University; and Inés Sanguino, Claire Shewbridge and Marco Kools of the OECD Secretariat. The OECD is grateful for the time and input of the DoET officials and all education stakeholders in the Flemish Community of Belgium interviewed in the preparation of this report.

OECD
BETTER POLICIES FOR BETTER LIVES

Table of Contents

1. Introduction

Background

The Flemish Community of Belgium has been among the strong performers on PISA during the last two decades, with 15-year-olds' mean scores in reading, mathematics and science consistently above the OECD average. However, student performance has been decreasing and more students fail to reach basic proficiency levels in these subject areas than before (OECD, 2019[1]). A similar decline in student performance is observed in primary education (Faddar et al., 2020[2]; Tielemans et al., 2017[3]). The declining results in large-scale international student assessments are corroborated by the Flemish National Assessment Programme (i.e. the "*peilingen*"). The decline in student performance has raised concerns about the school system's capacity to reverse the downward trend – or at least sustain current levels – and has created increasing pressure on the system to reconsider the conditions, incentives and support it provides to enable high-quality teaching and learning practices (OECD, 2021[4]).

Against this backdrop, the Flemish Community of Belgium has initiated a series of reforms in recent years to help students adapt to changing demands placed on their competences and strengthen the overall quality of education. These included updating the school curricula, with the introduction of new competency-based attainment targets or minimal learning objectives (OECD, 2021[4]). The new attainment targets have faced contestation from some education providers who considered that they impede the principle of "freedom of education", however. In addition to the curriculum reform, in 2019, the Flemish government announced the introduction of full cohort standardised student assessments in Dutch language and mathematics to assess the extent to which students are achieving attainment targets, as well as their individual and school's learning gains (Shewbridge and Köster, 2021[5]). Standardised student assessments are tests that are administered and marked in a systematic and consistent manner, whose validity and reliability have been empirically established (OECD, 2013[6]).

International research evidence has shown that many reforms and policy initiatives have failed to take hold in schools and classrooms, or at best, they get adopted on the surface without altering behaviours and beliefs. Evidence highlights that if, and how, key stakeholders are recognised and included in the design and implementation of a new policy or change initiative is crucial to its success (OECD, 2020[7]; 2020[8]; Viennet and Pont, 2017[9]; Burns and Köster, 2016[10]). A new policy or desired change is unlikely to succeed unless those expected to implement it see its value, want to see it happen, are confident in their capacity to implement it and ideally take ownership of it (McKnight and Glennie, 2019[11]; Schleicher, 2018[12]).

In 2020, the DoET invited the OECD to support the design and implementation of the standardised student assessment reform. The OECD Strategic Education Governance Project team began this work by consulting with a wide range of stakeholders on their motivations and concerns about introducing these assessments (Shewbridge and Köster, 2021[5]). The analysis confirmed that the Flemish education profession and many other education stakeholders were, in principle, supportive of the introduction of the full cohort standardised student assessments – signalling a notable shift in opinion and mindset compared to a previous OECD review on evaluation and assessment in education in 2011 (Shewbridge et al., 2011[13]).

There was some contestation of the reform plans, however, partially because of the initial uncertainty of what the reform would actually entail and how it would impact the daily work of schools, teachers and students. The analysis identified a number of lessons learnt or points of attention for designing and taking forward the standardised student assessments reform "with" the involvement and support of education stakeholders at different levels of the system (Box 1).

To support the DoET's efforts to take forward these points, the OECD Implementing Education Policies team was asked to support the DoET in successfully designing and implementing the standardised student assessment reform at primary and secondary levels. This report presents the analysis, key findings and recommendations to help realise this objective.

Box 1. Lessons learnt from the OECD's Strategic Education Governance Project case study

The Flemish Department of Education and Training invited the OECD to consult stakeholders on their motivations and concerns to introducing full cohort standardised student assessments in primary and secondary education (February-June 2021). The analysis identified several lessons or points of attention for further work by presenting stakeholder feedback and supporting evidence in six interrelated domains of a research-based strategic education governance framework.

Stakeholder involvement

Prioritising clear and active communication: The high-level forum[1] can serve as an authoritative communication channel and collect feedback from key stakeholders promptly and transparently. There is an opportunity to more actively involve stakeholders in the next stage of development, such as to provide input into clarifying the purpose(s) and uses of the standardised student assessments.

Committing to stakeholder involvement and ensuring key voices are heard: An important lesson is to take stakeholder involvement seriously at every stage of policy development. Mobilising awareness, support and feedback channels for school leaders will be critical. Supporting a student survey on their expectations of standardised assessments will empower student voice and provide pertinent insights.

Organising contributions from the educational field to support the Steunpunt:[2] There is motivation for being involved in test development and an opportunity to establish a coalition of test development partners across educational networks. The Steunpunt can facilitate this by providing clear guidance on scheduling and expected time commitments.

Strategic thinking and whole-of-system perspective

Developing, sharing and consolidating common goals and how standardised assessments will support these: The case study has found a shared concern on the overall quality of education in the Flemish Community of Belgium and a body of evidence to support this. Such widespread recognition is pivotal and presents an opportunity to create a common vision for the role of standardised student assessments. There is strong support for standardised student assessments as tools to support school quality development. It is important to consider safeguard measures to this effect, including ensuring schools are encouraged to continue to develop and innovate their practice.

Taking a long-term perspective and adapting to changing contexts and new knowledge: There is value in refining and evolving the development of standardised student assessments through concrete experiences in the educational field. This brings professional learning and development opportunities through collaboration between the research community (test developers) and schools. The first administrations of the standardised assessments will generate a lot of knowledge on optimising the use of results at the school level. An opportunity to clarify initial expectations is to ensure a coherent approach and communication from the Flemish Inspectorate of Education and the pedagogical advisory services (PDB) on how to use these results for school development as part of the broader view of educational quality (i.e. the Reference Framework for Quality in Education – OK Framework).

Co-ordinating action and learning from experiences in the educational field: There is value in providing co-ordinated guidance from the central authorities based on systematic input from the educational field, on the expected use of the standardised assessments, and the associated time and resource requirements for teachers and schools.

Capacity and knowledge governance

Ensuring technical capacity for standardised test development and administration: Strong credibility for the Steunpunt as a centre of scientific expertise will provide fertile ground for receiving regular feedback from the educational field during test development. A careful evaluation of schools' capacity to administer digital assessments needs to be performed and due attention paid to field trials.

Laying the foundations for the systematic use of standardised test results by professionals, with attention to:

- Skills – There is a need to give adequate attention to the capabilities of teachers and other school staff to work with the results of standardised and other assessments. There is an opportunity to commit to invest in professional development and in ways that can support collaborative practices in schools.
- Availability – The rapidity of results feedback will play into the perceived value and relevance of the results for educators. Notably, this would support students' expectations for the standardised assessments to bolster the culture of feedback to students on their progress more generally.
- Organisational processes – School leaders will drive the preparation of the necessary processes and structures to create the space for effective use of the standardised assessments. This can be supported at the system level by preparing common guidance material for schools – a process that will need to engage school leaders and teachers and mobilise the expertise of PDB.
- Interaction – The design and development of feedback from the standardised assessments will be strengthened by the direct interaction between researchers and schools. Importantly, this presents an opportunity to promote horizontal collaboration and learning across different educational networks.
- Standards – The development of guidance material for schools will provide a common anchor for expectations on using standardised assessments, clarifying how they relate to the existing central anchors of the attainment targets and the broader OK Framework. There are roles here for the Flemish Inspectorate of Education and the PDB to document expectations of how schools can interpret and position the data from the standardised assessments in a broader array of evidence.

Accountability

Ensuring the "fit" of accountability instruments: The case study notes the perception of "accountability" in Flemish education as a matter of internal responsibility and great resistance to the public availability of school performance information. There is an opportunity to place the standardised student assessments within the strengths of the current accountability system that focuses on dialogue and deepening the understanding between available data and links to ideas for improving practice.

Enhancing critical reflection on substantive expectations: Data from standardised assessments will provide an objective and external perspective for school development, with appropriate mechanisms for designing data use and interpretation by teachers and school leaders to support informed practice and strategic planning. The Inspectorate of Education can confer a valuable perspective to schools on how they use the results.

1. In 2021, the Flemish Minister of Education decided to establish a specific stakeholder consultation platform to facilitate communication and feedback at key stages of the development of standardised student assessments. The high-level forum was mandated to supervise key decisions in policy development for introducing standardised student assessments. The first meeting of this "high-level forum" was convened in May 2021.
2. The Steunpunt Central Tests in Education, or "Steunpunt" in short, is a consortium of all five Flemish universities and two universities of applied sciences that the Flemish government has commissioned to support the introduction of the standardised student assessment during the period 2021-25.

Source: Shewbridge and Köster (2021[5]), *Promoting Education Decision Makers' Use of Evidence in Flanders*, https://doi.org/10.1787/de604fde-en.

Overview of the project and methodology

The "Supporting the implementation of the standardised student assessment reform of the Flemish Community of Belgium" project aims to provide concrete recommendations for supporting the successful implementation of the standardised student assessment reform that the Flemish Community of Belgium has recently embarked upon. Building on the lessons learnt and recommendations offered through the OECD Strategic Education Governance Project (see Box 1), the OECD Implementing Education Policies team undertook an analysis of the design and implementation of the standardised student assessment reform.

Using the Implementing Education Policies' framework (OECD, 2020[8]) and the OECD framework on Evaluation and Assessment for Improving School Outcomes (OECD, 2013[6]) as lenses, this work consisted of a desk study of policy documents and studies and interviews and focus group discussions with a wide range of education stakeholders from different levels of the Flemish education system. Policy documents on the standardised student assessment reform were limited to a first concept note on the standardised student assessments that was published in January 2022, followed by a subsequent preliminary decree that was released later that year (July 2022) and a draft decree approved by the government in February 2023. At the time of finalising this report (March 2023), the decree was awaiting final approval by the Flemish Parliament. The decree is expected to enter into force in September 2023. In support of the analysis, the OECD team travelled to the Flemish Community of Belgium for a five-day visit in June 2022 to conduct interviews with a range of education stakeholders (see Annex A) and conducted a final online stakeholder event in June 2023.

The analysis allowed the OECD (see Annex B) to gain an in-depth understanding of the standardised student assessment reform and the broader education context and policy landscape, allowing it to formulate concrete guidance for enhancing the design and implementation of the standardised student assessment reform. These recommendations could be used to form the basis of a comprehensive implementation strategy.

Structure of the report

This report is structured in four sections. The remainder of Section 1 provides an overview of the Flemish school system that helps the reader cast the analysis presented in the following sections into context. Section 2 starts by looking into the governance of the standardised student assessments, including the overarching aim and purposes of the reform. This is followed by an examination of the procedures and methodologies associated with the assessments. The section concludes with a discussion on using the student assessment results. Section 3 focuses on the elements needed for a conducive context to support the successful implementation of the reform. These include stakeholder engagement, communication, clarity in roles and responsibilities, and the capacity of teachers and school leaders to effectively use the student assessment results for improvement. The last section focuses on developing an implementation strategy for the reform, including by providing concrete guidance for a monitoring, evaluation and research programme that could support the reform in and beyond its initial implementation.

School education in the Flemish Community of Belgium

Structure and governance of the Flemish school system

Structure of the school system

In the Flemish Community of Belgium, education is compulsory from age 5 to 18; nevertheless, as pre-primary education is generalised and free of charge, more than 97% of children attend (OECD, 2021[4]). From age 15 onwards, students may engage in part-time schooling and opt for a structured learning path which combines part-time vocational education in an educational institution with part-time employment. The school system is stratified, with a first streaming of students occurring at the beginning of secondary education. The school system is organised in four main stages:

1. **Pre-primary education** (typical ages: 2.5-6) is not mandatory, but over 90% of children are enrolled at age 3. Pre-primary and primary education is usually provided under the same roof in elementary schools.

2. **Primary education** (typical ages: 6-12) lasts for six years. There is also an offer of seven years of special primary education for children with special educational needs. At the end of primary education, students who achieve the objectives of the curriculum receive a certificate.
3. **The first stage of secondary education** (typical ages: 12-14) lasts for two years and is organised in two streams: the "A stream" and the "B stream". The vast majority of students are enrolled in the A stream. This first stage of secondary education is intended to provide students with a shared curriculum of basic general education. Students who did not receive a certificate for primary education and those who wish to pursue a vocational education enroll in the B stream.
4. **The second and third stages of secondary education** (typical ages: 14-18) usually last for two years each. Education is organised in different pathways depending on whether the student wants to enrol in tertiary education, vocational education or a dual path, including a mix of subjects with an academic and practical focus. For students in vocational secondary education, there is an option to take an additional year in the final stage of secondary education if they wish to enter tertiary education.

In the 2020/21 school year, there were 467 136 students in primary education and 476 200 in secondary education, including those in part-time vocational and special needs education (Flemish Department of Education and Training, 2022[14]).

Governance of the Flemish school system

Belgium has three tiers of government: the federal state, the regions and the communities. The federal government has responsibility for areas including social security, justice and defence. The jurisdiction of the three regions (the Flemish, Walloon and Brussels-Capital Regions) revolves mainly around territorial and economic matters. The three communities (the Flemish, French-speaking and German-speaking Communities) are responsible for matters related to the individual, including cultural, language and educational matters. The Flemish Region and the Flemish Community governments have merged into one. The Flemish Community is responsible for education in the Flemish Region and for education provided in Dutch as the main instructional language in the Brussels-Capital Region (Nusche et al., 2015[15]).

The Flemish, French- and German-speaking Communities each have autonomous education systems. Only a small number of education regulations remain at the federal government level. The Belgian Constitution states the federal government decides the dates of the beginning and end of the school year, the minimum requirements for diplomas, and teachers' pensions. However, even in those cases, the government heavily relies on consultations with non-governmental entities, such as teacher unions and education networks (OECD, 2018[16]).

Responsibilities for policy development and implementation

The Flemish Ministry of Education and Training is ultimately responsible for all stages of the educational system, and each school is responsible for providing a good quality education. The ministry comprises the DoET and three executive agencies. The DoET has responsibility for policy preparation, evaluation, co-ordination and communication, while three autonomous agencies are in charge of policy implementation and oversee all services related to quality improvement in education (Nusche et al., 2015[15]).

- **The Agency for Higher Education, Adult Education, Qualifications and Study Grants** (AHOVOKS) defines the minimum standards for quality education that all Flemish schools must meet. In consultation with education stakeholders, it sets schools' attainment targets and developmental objectives. It is also in charge of recognising qualifications and prior learning, including central tests for university admission in degrees such as medecine, dentistry and veterinary, and implementing policies on tertiary education, adult education and study grants.

- **The Agency for Educational Services** (Agentschap voor Onderwijsdiensten, AGODI) is responsible for implementing policies on school education, part-time arts education, centres for student guidance, the Inspectorate, and the pedagogical support to teachers and schools. The AGODI pays the salaries of all school staff; manages teachers' personal files; and monitors student enrolment, truancy and early school leaving.
- **The Agency for Educational Infrastructure** (Agentschap voor Infrastructuur in het Onderwijs, AGION) provides financial support for the acquisition, construction and renovation of buildings for grant-aided schools and universities.

School autonomy and freedom of education

The Flemish Community has one of the OECD's most devolved education systems, with schools enjoying a high degree of autonomy and the local level (provincial and municipal governments) playing only a minor role (OECD, 2018[16]). Flemish schools, in fact, benefit from the highest levels of autonomy among OECD countries in all aspects of education, including assessment practices (OECD, 2020[17]). School autonomy is grounded in the constitutionally guaranteed principle of "freedom of education", which gives any natural or legal person the right to set up a school, recruit staff and determine the (educational, religious or ideological) principles of the school.

Parents are also free to choose and are guaranteed access to the school of their choice within a reasonable distance from their home. In addition, funding "follows the student" as long as the school complies with a number of requirements. These include, among others, accepting the educational structure imposed by decree, participating in educational inspections, and following a curriculum that involves attaining targets and development goals. For schools to receive public funding and have the right to award official certificates, they must also be "recognised" by the Flemish authorities. There is a system of compulsory inspection for all schools seeking recognition by the Flemish government. Non-recognised schools constitute less than 1% of the provision (Shewbridge and Köster, 2021[5]; Nusche et al., 2015[15]).

A previous OECD review noted that usually, *"in exchange for increased autonomy, schools [in OECD countries] face increased accountability, so that parents can make their choices based on information about school quality and performance. The Flemish Community does not have as much overt accountability nor does it have as many diverse forms of accountability. […] Based on experiences in other countries, one might expect demands for greater transparency and more direct access to data on schools, but such features are not characteristic of the Flemish approach*" (Nusche et al., 2015, pp. 103-104[15]). The freedom of education thus makes the Flemish system one of the most devolved systems in the OECD and also a system in which the accountability mechanisms of schools towards the central government are not strongly developed. In such a context, steering by the centre becomes more difficult, and risks of system imbalances are higher (Rouw et al., 2016[18]) (see below).

The role of school boards and educational networks

All schools in the Flemish Community are governed by a school board, which can be responsible for one or more schools. These legally recognised bodies oversee the implementation of legislation. They have wide autonomy and can decide freely on their teaching methods, curricula, timetables and staff appointments. The government sets conditions only for recognising a school and granting financing (Eurydice, 2022[19]). There are about 1 500 school boards in the Flemish Community. Often several school boards within the same network are grouped under an "umbrella organisation", representing them in policy discussions with the government and providing school support, such as by developing curricula and timetables based on the centrally defined attainment targets and developmental objectives.

Officially recognised schooling in the Flemish Community is organised within three educational networks (see Table 1 and Figure 1):

- **The Flemish Community education network** (Onderwijs van de Vlaamse Gemeenschap, GO!) has an autonomous central board, the Council of the GO!, that acts as a single provider on behalf of the Flemish Community. The Council of the GO! is not responsible for the organisation of schools, as this is a devolved competence to the school groups, which have their own chair and general director.
- **The publicly funded and publicly managed education network** (Officieel gesubsidieerd onderwijs, OGO), also referred to as grant-aided public education, includes schools organised by the provincial and city/municipal authorities. These local and provincial authorities act as school boards. At the political level, the city and municipal authorities are represented by the Educational Secretariat of the Association of Flemish Cities and Municipalities (Onderwijssecretariaat voor Steden en Gemeenten van de Vlaamse Gemeenschap, OVSG). The Flemish Provincial Education (Provinciaal Onderwijs Vlaanderen, POV) represents the provincial authorities.
- **The publicly funded and privately managed education network** (Vrij gesubsidieerd onderwijs, VGO), also referred to as grant-aided private education, includes denominational and non-denominational schools. The vast majority of denominational schools are of Catholic tradition. The school boards of Catholic schools are typically private foundations related to dioceses, parishes or congregations. These are represented by the Catholic Education Flanders (Katholiek Onderwijs Vlaanderen, KOV). Six other denominational schools are clustered in the Council of School Boards of Protestant-Christian Education (IPCO). The few other denominational schools have not established an umbrella organisation. Non-denominational schools typically pursue a particular educational method or are based on a particular philosophy. These schools are represented by the Federation of Steiner Schools or the Federation of Independent Pluralistic Emancipatory Method Schools (FOPEM). There are also those schools clustered in the Flemish Education Consultation Platform (Vlaams Onderwijs Overleg Platform, VOOP) that do not pursue a particular educational method but are based on liberal humanism. The four smaller umbrella organisations within the VGO network have established the Consultation Body of Small Education Providers (Overleg Kleine Onderwijsverstrekkers, OKO) as a discussion partner for the Flemish Community (Nusche et al., 2015[15]).

Table 1. Educational networks, umbrella organisations and school boards in the Flemish Community

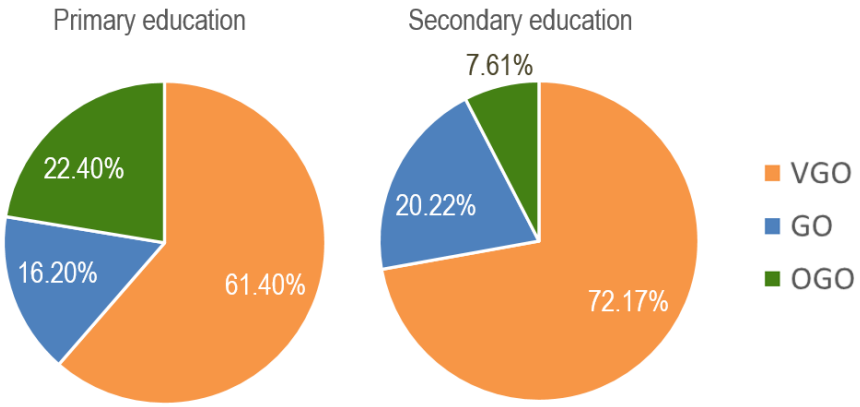
<table>
<tr><td>Network</td><td>Community education (GO!)</td><td colspan="2">Grant-aided public education (OGO)</td><td colspan="3">Grant-aided private education (VGO)</td></tr>
<tr><td rowspan="2">Umbrella organisation</td><td rowspan="2">Flemish Community Education (GO!)</td><td rowspan="2">Educational Secretariat of Flemish Cities and Municipalities (OVSG)</td><td rowspan="2">Flemish Provincial Education (POV)</td><td rowspan="2">Catholic Education Flanders (KOV)</td><td>Council of School Boards of Protestant-Christian Education (IPCO)</td><td>FOPEM Steiner, VOOP</td></tr>
<tr><td colspan="2">» Consultation Body of Small Education Providers (OKO)</td></tr>
<tr><td>School board</td><td>Flemish Community Education Council
» School groups</td><td>Cities and municipalities</td><td>Provinces</td><td colspan="3">Private foundations</td></tr>
<tr><td colspan="7">Schools</td></tr>
</table>

Note: FOPEM = Federation of Independent Pluralist Emancipatory Method Schools; VOOP = Flemish Education Consultation Platform.

Source: Adapted from OECD (2015[20]), *School education in the Flemish Community of Belgium*, https://doi.org/10.1787/9789264247598-5-en.

.

Figure 1. Distribution of students across the three largest educational networks in the Flemish Community in the 2020-21 academic year



Notes: VGO = grant-aided private education; GO = Flemish Community education; OGO = grant-aided public education. Data for the 2020/21 academic year. Secondary schools only include data from full-time regular schools, excluding special secondary education and part-time vocational secondary education.
Source: Based on Flemish Department of Education and Training (2022[14]), *Flemish education in figures 2020-2021*, https://publicaties.vlaanderen.be/view-file/48569.

Curriculum, quality assurance and standardised student assessments

Curriculum – attainment targets and educational goals

The Flemish Parliament approves a "core curriculum" consisting of attainment targets and developmental objectives to be implemented by all schools, except those with equivalent attainment targets approved, which are currently only Steiner schools. Final attainment targets are minimum learning objectives the government considers necessary and attainable for students at the end of specific year levels and study programmes. One example of an attainment target for primary education in mathematics is "*The students can count and count back with units, pairs, quintuples and powers of ten*" (AHOVOKS, n.d.[21]). They were first implemented in 1998 to increase transparency, quality assurance and comparability in the education offered across schools and networks. Developmental objectives are goals that schools should strive for their students to attain, but there is no obligation for all students to reach them. Schools only need to account for their efforts to work towards these goals. The attainment targets are an instrument for the government to guarantee the minimum desired quality of education. Within the framework of attainment targets and developmental objectives, schools are free to develop their own curricula, which will reflect different priorities and cover broader areas, but must allow the Flemish authorities to quality assure the schools via regular inspections (see below). Most schools, however, work within the curricula and supporting tools (e.g. learning plans) developed by the umbrella organisations of their educational network. Flemish schools, as noted, are affiliated with umbrella organisations which offer curricula and assessment support via their "pedagogical advisory services" (see below). Schools can choose whether to use these services, including standardised student assessments (Nusche et al., 2015[15]).

One of the criticisms of the attainment targets has been that they lacked clarity and were not appropriately aligned across educational stages (i.e. pre-primary, primary and secondary education) (Rouw et al., 2016[18]). These and other challenges have contributed to the efforts to reform the primary and secondary attainment targets in recent years (see below).

School-based curricula and internal quality assurance

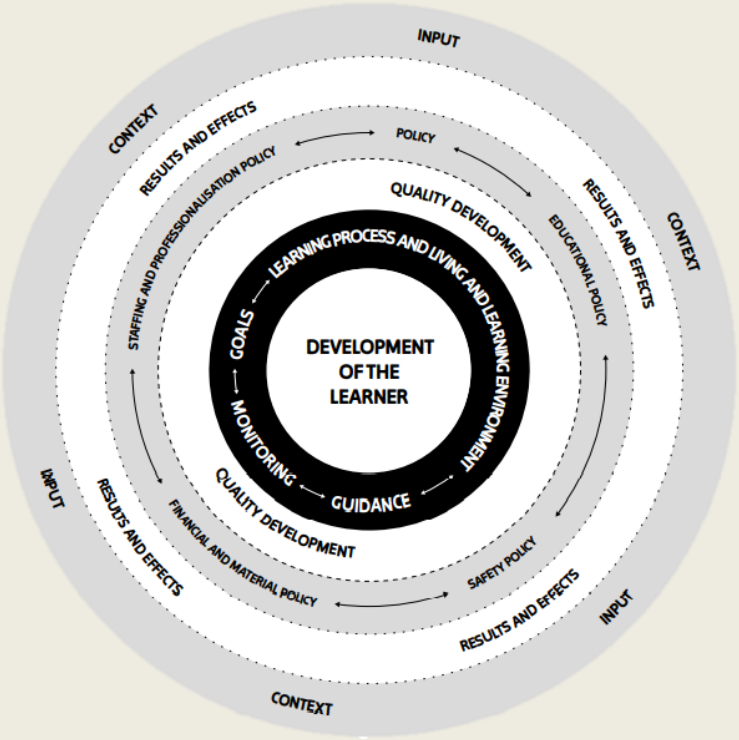
According to the 2009 Decree on Quality of Education, each school is responsible for providing a high-quality education. As mentioned above, within the framework of attainment targets and developmental objectives, schools are free to develop their own curricula and assessment practices. Each school is responsible and free to decide how to internally monitor education quality and student learning. The former is based on the expectation set in the Reference Framework for Quality in Education, hereafter referred to as the "OK Framework" (Flemish Ministry of Education and Training, 2018[22]) (see Box 2). OK is the Dutch acronym for educational quality. The OK Framework was introduced in 2018 and consists of 37 indicators of educational quality. Schools can use this framework of quality indicators in the self-evaluation process, but importantly, it also guides the work of the Inspectorate of Education when externally assessing the quality of schools. It also informs the work of the PDB – this collaboration is often referred to in the Flemish Community as the "quality triangle approach" (see Figure 3). In addition, Flemish schools have access to a wide range of other resources to support them in their quality assurance and school development processes, offered to them by the DoET, the Inspectorate of Education, school advisory services and other partners.

Box 2. The Flemish Reference Framework for Quality in Education

In 2018, the Flemish Community introduced the Reference Framework for Quality in Education, or "OK Framework", to support schools in their self-evaluation and development planning. It was developed through a combination of research and a consultative process with a broad range of stakeholders (e.g. students, parents, teachers and school leaders, pedagogical advisors, teacher trainers, education inspectors, experts, and trade unions). It sets the expectations of good quality education for all levels, excluding tertiary education, while respecting schools' autonomy.

The framework consists of 37 indicators of educational quality that are part of 4 interrelated overarching categories that interact with each other: 1) results and effects; 2) stimulating development; 3) quality development; and 4) policy. As depicted in Figure 2, the development of the learner is at the centre of the framework.

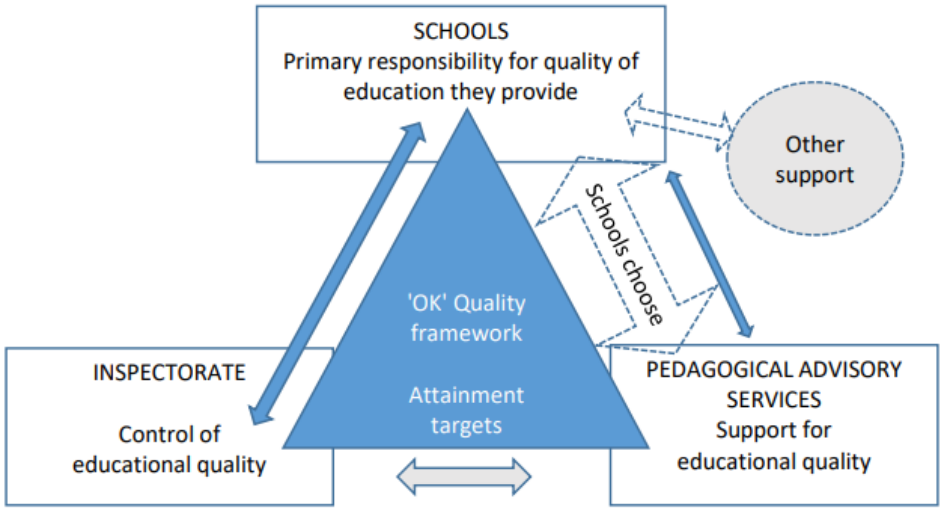
Figure 2. Overview of the Reference Framework for Quality in Education (OK Framework)



The framework has helped create a shared vision or common language as to "what is a good school?" in the Flemish context. It is used by schools for their quality assurance processes, as well as by the Inspectorate of Education and the pedagogical advisory services (the "quality triangle approach", see Figure 3). The framework can be considered an example of an initiative incentivising the building of trust and creating ownership of schools and teachers of a system-wide, systemic evaluation practice (Shewbridge, Fuster and Rouw, 2019[23]). The framework has only been implemented since 2018. Further work remains to be done to promote and fully embed its use in all Flemish schools' quality assurance and developmental processes; it is widely believed to be an important policy instrument for supporting schools in their improvement efforts.

Source: OK Education Quality (2016[24])*The reference framework for Quality in Education: quality expectations and quality images,* *https://www.onderwijsinspectie.be/sites/default/files/2022-06/OK_magazine_eng.pdf*.

Figure 3. The "quality triangle approach" to educational quality in the Flemish Community of Belgium



Notes: At the top of the triangle are autonomous schools, legally responsible for the quality of the education they provide. Elements in blue are underpinned by regulation. Within the main triangle, schools are legally required to help ensure their students achieve the attainment targets. Schools are obliged to receive an inspection from the Inspectorate. The arrow is long and narrow, indicating the lengthy school inspection cycle. The Inspectorate of Education can obligate a school to engage external support to improve the quality of its educational offer. However, in reality, this is not that frequently implemented; in 8.1% of the schools (i.e. 27 schools) that were inspected in the 2019/20 school year and 5.5% of the schools (i.e. 29 schools) in the 2018/19 school year. All umbrella organisations, except small school providers, offer educational support via pedagogical advisory services (PBD). Schools can choose to use the curriculum developed by the PBD and also to use standardised student assessments, currently developed by one PBD for Grades 4 and 6 and another PBD only for Grade 6. Many schools choose the support from the PBD. The narrow blue arrow, therefore, indicates stronger and closer connections, although these remain trust-based and voluntary. Representatives from the Inspectorate and the PBD have reported that there are strengthened professional connections between their respective roles since the co-construction of the OK Framework. For further details on the "quality triangle approach" to educational improvement, see Shewbridge and Köster (2021[5]).

Source: Shewbridge and Köster, (2021[5]), *Promoting Education Decision Makers' Use of Evidence in Flanders*, https://doi.org/10.1787/de604fde-en.

External quality assurance – The Inspectorate of Education

As in many OECD countries, the Inspectorate of Education is the major mechanism to hold schools in the Flemish Community to account (OECD, 2013[6]). The Inspectorate of Education is an independent body that evaluates whether schools adhere to regulations and achieve minimum standards around quality and processes in place (attainment targets), using a long inspection cycle. The intention is to conduct an inspection of each school every six years, but in reality, and partly because of the pandemic and reforms, the cycle can result in one inspection every ten years or more. The Inspectorate varies the frequency and intensity of visits based on factors such as the institutional profile and previous assessment reports. Nonetheless, ten years is a long time compared to other OECD countries and jurisdictions (OECD, 2015[25]).

Evidence from school inspections in recent years shows considerable scope for strengthening schools' quality assurance and development processes both at primary and secondary levels. For example, in four out of ten schools inspected in 2019/20, the Inspectorate found that systematic self-evaluation of quality did not meet expectations, as they were generally fragmentary and not school-wide (Dutch Inspectorate of Education, 2021[26]).

Finally, it is believed that the recently developed OK Framework will benefit both internal and external quality assurance. A clear challenge for the Inspectorate's system-level and school-level monitoring of education quality has been the absence of a full cohort standardised student assessments. The Inspectorate sees the regular school-level data that the standardised student assessment reform will provide as an important opportunity to augment its information base for school inspections and implement a more differentiated approach to inspections.

Central examinations

Within OECD countries, central examinations are widely used to certify student learning and, in some circumstances, for quality assurance and supporting improvements at different levels of the system, including schools (OECD, 2013[6]). Although central examinations are rarely used in the early years of schooling, they become more widespread at the higher levels (e.g. Abitur in Germany, Baccalauréat in France, end-of-primary test in the Netherlands, and Matura in countries such as Poland or Slovenia). In the Flemish Community, however, there are no central examinations for students at the end of compulsory education – an approach shared by only a handful of OECD countries (Rouw et al., 2016[18]).

Standardised student assessments

The use of standardised student assessments in primary and secondary education is commonplace in OECD countries and over the past 20 years the majority of OECD countries have chosen to administer full cohort assessments; that is, with all students and schools participating (Shewbridge and Köster, 2021[5]). Depending on their purpose(s), these assessments can help governments, education authorities and schools to see whether students are reaching attainment targets or learning outcomes, identify misconceptions and areas for improvement, and develop remediation suggestions. They can provide valuable information about how education programmes are working, allow for identifying and examining strengths and/or "good practices" that can be shared across the system, while importantly also pointing to areas for improvement and schools needing support (OECD, 2013[6]; Shiel, Kellaghan and Moran, 2010[27]; Jackson, Adams and Turner, 2017[28]; DeMatthews, Knight and Woulfin, 2021[29]).

There have been no full cohort standardised student assessments to measure the learning outcomes of all students at key stages of schooling in the Flemish Community – an exception among OECD countries (Shewbridge and Köster, 2021[5]). A situation that is to change with the ongoing standardised student assessment reform. However, there is a range of sample-based assessments whose results schools can use to inform their quality assurance processes. These include the National Assessment Programme, introduced in 2002 as a periodical sample-based assessment to monitor the implementation of attainment targets at the system level. These student assessments, called the "*peilingen*" in Dutch, are administered every year to a representative sample of 1 500-3 000 students in primary school (Grade 6) and secondary school (Grades 8, 10 or 12) in different learning areas based on the expected level of learning, as specified in the central attainment targets. Each assessment cycle examines a different learning area, such as French, Dutch or biology, on a rotating basis. All participating schools receive a feedback report of their own performance to support them in their internal quality assurance processes. All Flemish schools can also voluntarily choose to administer a parallel version of these central assessments, called the "*paralleltoets*" in Dutch, to support their internal quality assurance. Few schools, however, make use of the option. Of all full-time schools (excluding special education), 209 primary schools (8% of the total) and 30 secondary schools (3% of the total) took part in a "*paralleltoets*" in 2022.

In addition, the umbrella organisations of several educational networks offer student assessments for the final year of primary education. These assessments are aligned with the networks' respective curricula. KOV and OVSG provide the two main assessments. The KOV assessments (Inter-Diocesane Proeven, IDP) test students' proficiency in Dutch language, mathematics, French, arts education, science and

technical education, and "humanity and society"). These assessments are taken by almost 90% of all Flemish Catholic school students. Results are reported compared to the average of participating schools and participating schools with a similar student population or contextual characteristics.

The OVSG assessments (in Dutch called the OVSG-toets) are used by schools managed by the municipalities and cities but also by most schools of the GO! network and some publicly funded private schools. These assessments cover the breadth of the OVSG curriculum and attainment targets in Dutch language, mathematics, science and technical education, humanity and society, arts education, and French. Practical assessments are also available in a range of subjects, including spoken language (Dutch and French), physical education, technology and music. Results are processed on line and can be compared to average results of participating schools, schools with a similar profile and previous years' results.

Partly in response to the growing concerns about the quality of education, since the 2017/18 school year, all students in the Flemish Community of Belgium at the end of primary education were to take a validated student assessment in at least two areas of study. In 2018/19, this requirement was expanded to a minimum of three study areas. Schools can choose to administer any of the above-mentioned Flemish sample-based student assessments.

Accountability

Results-based accountability remains rather weak in the Flemish school education system, including for school boards and educational networks. The Inspectorate has no mandate to evaluate school boards. There are neither central examinations at the end of compulsory education nor full cohort standardised student assessments that could give a comparable insight into students' performance against national attainment targets, thereby limiting their use for provider- and system-level monitoring. This makes the Flemish school system an outlier among OECD countries.

To provide information on student outcomes at the system level, the Flemish Community relies on a sample-based National Assessment Programme (i.e. the "*peilingen*") and its participation in large-scale international assessments (on three-, four- or five-year cycles). While it is not compulsory for schools to participate in the National Assessment Programme, response rates are generally high.

Schools are required to have a quality assurance system in place but are free to design it. Schools' internal quality assurance processes remain variable and, in many cases, underdeveloped (Shewbridge and Köster, 2021[5]). Schools can use standardised student assessments that are provided by the government or by the PDB in their respective networks. At the end of primary education, schools must administer a student assessment (in at least three subjects) that they choose from a validated and approved set. However, given the variety of student assessments used, a standardised comparative measure of student performance across all schools in the Flemish Community is lacking. Parents can consult schools' inspection reports; however, these are not always easy to understand, and some are outdated. Parents in the Flemish Community of Belgium do not have other sources to assess the educational quality of the school their children attend or plan to attend.

In sum, although the use of different accountability measures and tools varies across countries, from an internationally comparative perspective, the evidence points to an underdeveloped results-based accountability culture in the Flemish school system.

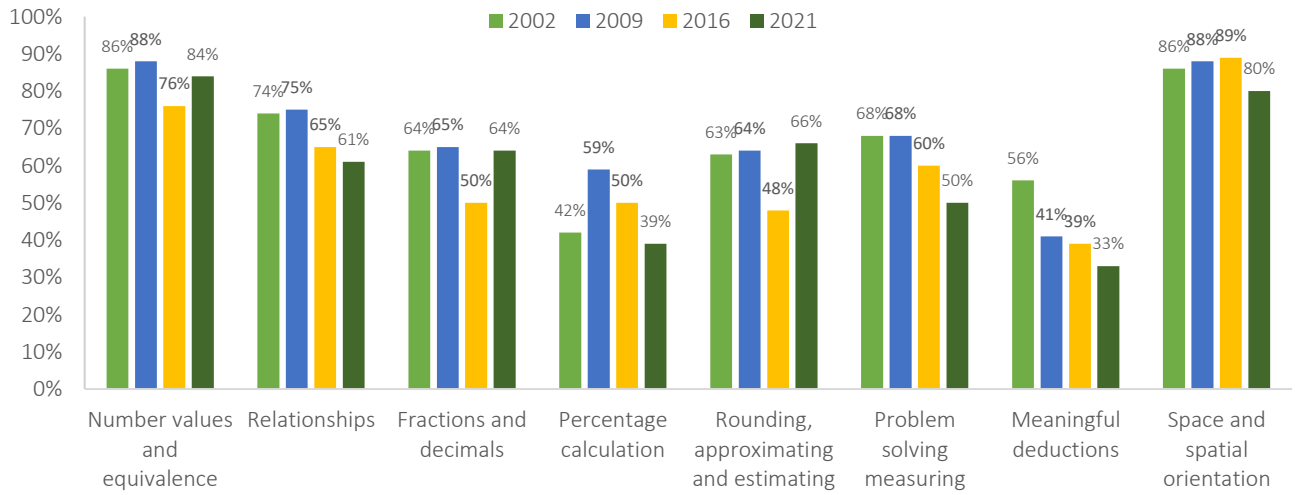
Quality and equity of schooling

There is a shared concern among Flemish education stakeholders about the overall quality of education. The debate about this has been informed, among others, by the collective dialogues on the results from the sample-based National Assessment Programme (i.e. the "*peilingen*"), as well as by the results from several international assessments such as the OECD's Programme for International Student Assessment

(PISA), the International Education Agency's (IEA) Progress in International Reading Literacy Study (PIRLS), and Trends in International Mathematics and Science Study (TIMSS). These standardised student assessments generally point to declining results, with significant equity challenges (Faddar et al., 2020[2]; Shewbridge and Köster, 2021[5]; Tielemans et al., 2017[3]; OECD, 2019[1]).

For example, at the primary level, the 2021 National Assessment Programme shows that a considerable proportion of students in Grade 6 do not meet the mathematics attainment targets (see Figure 4). Certainly, 2021 results should be interpreted with caution given the likely negative impact of the COVID-19 pandemic on students' learning and the educational sector more broadly. However, 2016 mathematics outcomes also depict a declining trend in results from previous assessment cycles in most attainment targets, with 50% or fewer students meeting the attainment targets in four of the learning areas. Significant student performance differences exist between learning areas (see Figure 4). In the Dutch language assessment, the percentage of students reaching writing learning targets ranged from 48% to 77%, depending on the learning area. From 2013 to 2018, there was a significant performance decline in reading (84% compared to a previous 92%) and listening (82% compared to a previous 87%) (Flemish Ministry of Education and Training, 2018[30]).

Figure 4. Percentage of students assessed in Grade 6 who met the mathematics attainment targets across learning areas

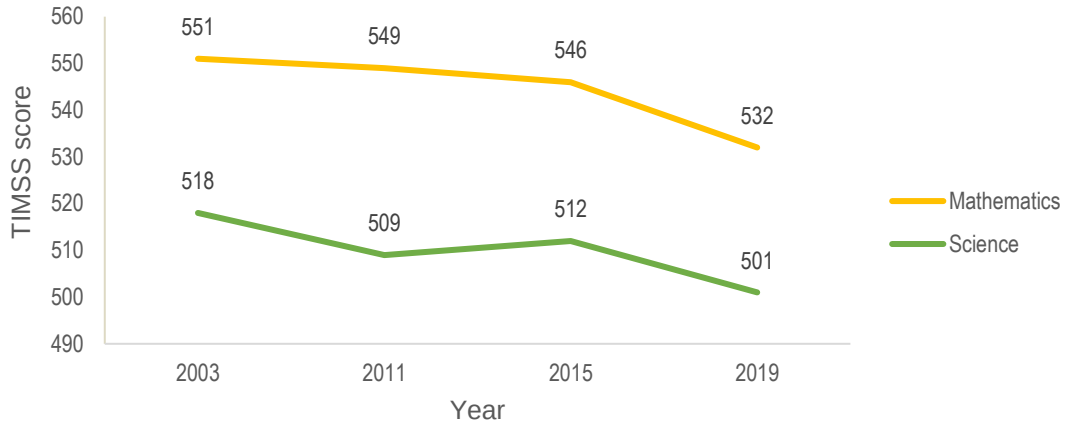


Note: Data provided by the Flemish Department of Education and Training.

The data also point to significant equity challenges. In mathematics, the percentage of those students with low socio-economic status who reach the attainment targets is sometimes half that of their peers with a high socio-economic status (22% to 57% respectively) (Flemish Ministry of Education and Training, 2022[31]). There are also significant performance differences between students with different socio-economic statuses in Dutch. While this performance gap narrowed in listening between 2013 and 2018, it has remained the same for reading (Flemish Ministry of Education and Training, 2018[30]). The Flemish Community aims to carefully examine the impact of the COVID-19 pandemic, as this may have had a greater impact on socio-economically disadvantaged students (Azevedo et al., 2022[32]).

Students' results in international student assessments align with those of the National Assessment Programme. In TIMSS, students' average performance has decreased in every cycle since 2003 in both mathematics and science (see Figure 5). In the case of reading comprehension, the most recent published PIRLS results indicate a significant decline from 2006 (547 points) to 2016 (525 points) (Mullis et al., 2017[33]).

Figure 5. Changes in average performance in TIMSS – Grade 4



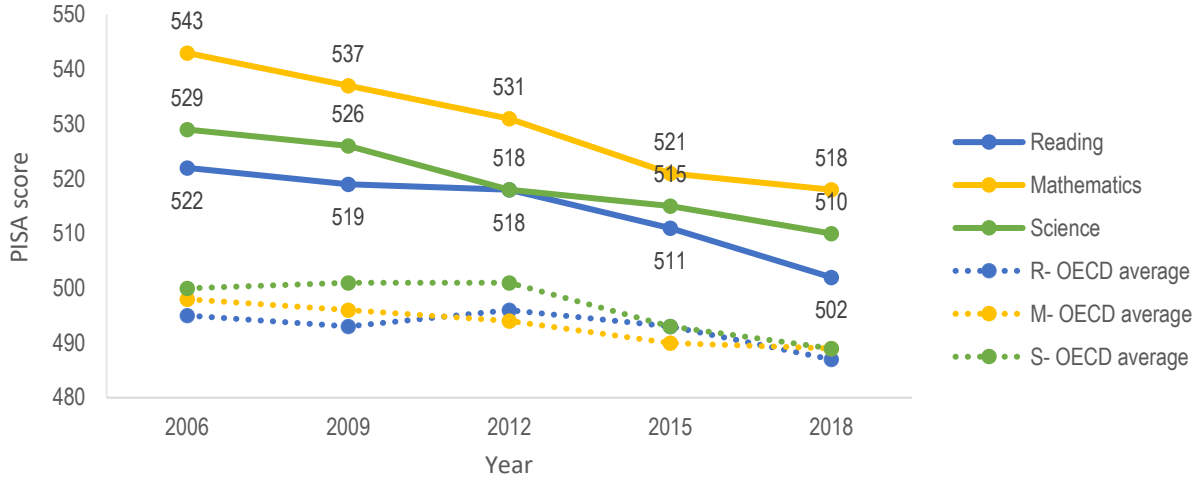
Source: Adapted from Shewbridge and Köster (2021[5]), *Promoting Education Decision Makers' Use of Evidence in The Flemish Community of Belgium*, https://doi.org/10.1787/de604fde-en. Data from Mullis et al. (2020[34]), *TIMSS 2019 International Results in Mathematics and Science*.

At the secondary level, results from the National Assessment Programme also show that a large proportion of students fail to meet the attainment targets in mathematics. In some instances, less than 30% of students achieve the targets in Grade 8 (in the A stream, the general education stream in secondary education). Regarding students in the vocational track (i.e. the B stream), 2021 results show a decline in the proportion of Grade 12 students achieving the attainment targets compared to 2013 in functional numeracy (26% compared to 39%), functional listening skills (30% compared to 39%) and information processing (53% compared to 62%). There was no significant difference in reading skills (34% compared to 38%). Although the COVID-19 pandemic may, at least partially, explain the decline in student performance, more than half of the students in 2013 were already failing to meet the attainment targets.

Similar to the student assessments in primary education, in some learning areas (e.g. algebra), the proportion of secondary students with a high socio-economic status that reaches the attainment targets is twice as high as those with a low socio-economic status (Flemish Ministry of Education and Training, 2019[35]; 2018[36]). Students who speak Dutch exclusively at home perform significantly better in most learning areas in both streams.

When looking at the results of international student assessments at the secondary level, 15-year-old Flemish students have consistently performed above the OECD average in reading, science and mathematics in PISA. However, student performance has been decreasing (see Figure 6), with more students failing to reach basic proficiency levels in these subject areas than before (OECD, 2019[1]). PISA 2018 showed that students' socio-economic status and immigrant background were strong predictors of student performance in the Flemish Community of Belgium (OECD, 2019[37]). There was a 110-point difference between the mean score of students in the top and those in the bottom quarter of socio-economic status (compared to an OECD average of an 89-point difference). Socio-economic status explained 17% of the variance of students' performance in reading, compared to an OECD average of 12% (OECD, 2019[37]). Only in three other OECD countries (out of the current 38) was this percentage higher (France, Hungary and the Slovak Republic). PISA 2018 showed that students' socio-economic status was even a stronger predictor of their performance in mathematics and science. These findings point not only to challenges in terms of the quality of education in Flemish schools – but also in terms of equity in education.

Figure 6. PISA results for the Flemish Community of Belgium, 2006-18



Note: The data for this figure were collected before Costa Rica became an OECD member in 2020. In 2015, there were changes to the test design, administration and scaling of PISA. See the Reader's Guide and Annex A5 of PISA 2015 Results (Volume I): Excellence and Equity in Education OECD (2016[38]) for a discussion of these changes.

Sources: OECD (2019[1]), *PISA 2018 Results (Volume I): What Students Know and Can Do, https://dx.doi.org/10.1787/5f07c754-en*; OECD (2016[38]), *PISA 2015 Results (Volume I): Excellence and Equity in Education*, https://dx.doi.org/10.1787/9789264266490-en; OECD (2014[39]), *PISA 2012 Results: What Students Know and Can Do (Volume I, Revised edition, February 2014): Student Performance in Mathematics, Reading and Science, https://dx.doi.org/10.1787/9789264208780-en*; OECD (2010[40]), *PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I)*, https://dx.doi.org/10.1787/9789264091450-en.

Policy priorities and recent developments

The Flemish Community of Belgium has initiated a number of new reforms and policy initiatives in recent years. These include measures to provide additional guidance for gifted students; a language integration programme for those with limited Dutch language skills; new primary education admission requirements; efforts to tackle teacher shortages; the establishment of new minimum attainment targets; and the introduction of the full cohort standardised student assessments. This section elaborates on the three latter.

Tackling teacher shortages

Although a shortage of teachers is not new to the Flemish Community of Belgium, concerns about shortages are growing and evidenced by the fact that teacher vacancies have doubled compared to five years ago (Growling, 2022[41]). Teacher shortages vary between regions, but are most acute in East Flanders, Antwerp and Brussels. In primary education, the shortage peak is still to come. The demand for teachers will rise sharply from 2025 and it is predicted that by 2026 it will exceed 2 000 full-time equivalents per year. That figure has been between 1 500 and 2 000 annually in recent years. Secondary education is seeing a large increase in the student population at present. As a result, a recruitment need of more than 4 700 full-time equivalents was forecasted for the school year 2022/23. The demand for teachers in secondary education is expected to decrease again from 2025 onwards (Flemish Department of Education and Training, 2019[42]).

In response, the Flemish government has made one of its strategic objectives to "*motivate and keep quality education professionals in education*" (Flemish Parliament, 2021[43]). A package of measures such as changing the seniority system, uplifting fiscal constraints for retired teachers who want to continue

teaching, ensuring teachers receive remuneration for extra teaching hours and providing laptops for all teachers were introduced at the start of the 2022/23 school year. Other measures include a more practical component of teachers' tertiary education thesis, an option to gradually return from sick leave and creating a teachers' (help) desk that guides people interested in becoming a teacher. These and other measures are aimed at attracting more people to the education profession in the coming years (Flemish Ministry of Education and Training, 2021[44]).

New attainment targets

Attainment targets are defined as minimum learning objectives which the parliament (and society more broadly) expects students to attain at different levels of education. In 2018, the Flemish government approved the new attainment targets for the first stage of secondary education (Flemish Parliament, 2018[45]), followed in 2021 by the approval of those for the second and third stages of secondary education (Flemish Parliament, 2021[46]). The new attainment targets have been contested, however, with some stakeholders raising concerns about their number and level of detail contradicting the constitutional freedom of education. Belgium's Constitutional Court dictated in June 2022 that the new secondary attainment targets violated the constitutional right to freedom of education and, therefore, were annulled (Constitutional Court, 2022[47]). Given that implementing changes takes time and that part of the new targets had already been introduced for the 2021/22 school year, the court provided a transitional phase in which schools may continue to use them until the 2024/25 school year.

The current plan is to amend the attainment targets for Stages 2 and 3 in secondary education which can take effect from September 2023. Afterwards, the attainment targets for Stage 1 will be reconsidered and aligned with the targets for Stages 2 and 3. The development of the new primary attainment targets has been put on hold.

Introduction of full cohort standardised student assessments in Dutch language and mathematics

In 2019, the Flemish government decided to introduce full cohort standardised student assessments in primary and secondary education. In 2021, the first government concept note on the standardised student assessments was released, followed by a subsequent preliminary decree (in July 2022). In February 2023, the government approved the final version of a draft decree. At the time of finalising this report (March/April 2023), the decree was awaiting final approval by the Flemish Parliament. The decree is expected to enter into force in September 2023.

The draft decree states that students are expected to take the student assessments in Grades 4 and 6 at the primary level and in Grades 8 and 12 at the secondary level. The standardised student assessments serve the overarching aim of strengthening and monitoring the quality of education through the achievement of attainment targets and by measuring the learning gain at the levels, as indicated below:

- At the system level, the extent to which the attainment targets are achieved, the extent to which learning gains are generated and as an element of quality assurance at the system level.
- At the school level, as an element of internal and external quality assurance:
    - As one of the elements in the internal quality assurance of the school. To this end, the school and the school board will be able to consult their own feedback report with contextualised results in a secure manner.
    - As one of the elements in the functioning of the PDB.
    - As one of the elements in the functioning of the Inspectorate of Education.

- At the level of the student group or class, as one of the elements for reflection on the pedagogical-didactic action in the school team.
- At the student level, as one of the possible elements that the class council can consider in the evaluation. Student-level results are not used as the sole criterion for evaluation (Flemish Government, 2022[48]).

The multiple purposes of the standardised student assessment are further discussed in Section 2.

The assessments will focus on Dutch (reading, writing and grammar) and mathematics (Flemish Government, 2022[48]). Mathematics and Dutch reading comprehension will be designed as computer-based, adaptive assessments. The Dutch writing assessment will be on line, but not adaptive. Computerised adaptive assessments are a form of computer-based testing in which the level of difficulty of the assessment questions adapts to the student's ability. This is done by automatically selecting question items based on the student's response to a previous (set of) question(s) (National Council on Measurement in Education, 2019[49]). The assessments will be standardised (i.e. administered and scored in a consistent manner such that standards are maintained over time); the intention is to administer them at schools according to a procedure that is the same for all.

The assessments will also be as inclusive as possible. Reasonable accommodations will be organised as much as possible for students who need them. Students entitled to adaptations or special educational materials during the school year may retain them for the assessments. Students in special education are not obliged to participate, with the exception of students in training path 4 of secondary education ("buso OV4"). However, these students can also be exempt from the decision of the school council for a particular reason (Flemish Government, 2022[48]).

The Steunpunt Central Tests in Education, or "Steunpunt" in short, is a consortium of all five Flemish universities and two universities of applied sciences that the Flemish government has commissioned to support the introduction of the standardised student assessment during the period 2021-25. Their main tasks consist of developing the assessment items, analysing assessment results, designing the feedback and supporting schools with the assessment administration, as well as using the feedback to improve the education quality inside the school. It unites researchers with the right expertise to draw up valid, reliable and transparent assessments, in consultation with the educational field, with the objective to start introducing the assessments in the 2023/24 school year (see Table 2).

The draft decree includes the intent for feedback at four levels: at the system-, school-, class- and student levels. At the system level, the results will be published in reports. These results serve the purpose of system monitoring.

At the school level, the school feedback report (i.e. dashboard) will contain contextualised performance data taking into account characteristics of the school's student population (such as the home language, the mother's educational level, etc.). In this way, schools can receive information that can be used for quality assurance and improvement purposes.

Table 2. Timeline for the introduction of Flemish assessments in primary and secondary education

School year	Primary education	Secondary education
2022/23	Calibration (preliminary study) in Grade 4	Calibration (preliminary study) in Grade 8
2023/24	First administration in Grade 4	First administration in Grade 8
2024/25	Calibration (preliminary study) in Grade 6	
2025/26	First administration in Grade 6	Calibration (preliminary study) in Grade 12
2026/27		First administration in Grade 12

Source: Flanders is Education and Training (n.d.[50]), Policy. Education regulations in development. Flemish tests, https://onderwijs.vlaanderen.be/nl/beleid/onderwijsregelgeving-in-ontwikkeling/vlaamse-toetsen. Accessed on the 31 August 2022.

At the classroom level, it is planned to provide teachers with rich information about the extent to which the class achieves the attainment targets for mathematics and Dutch. The feedback is to include information about strengths and areas for improvement for the class and is aimed to support reflection on pedagogical-didactic actions by the school team.

Lastly, schools will receive a feedback report for each student, indicating the student's level of proficiency, as well as the extent to which she/he has achieved the attainment targets. There will be no assessment in the form of pass or fail (Flemish Government, 2022[48]); there will be a scale score.

2. Towards a comprehensive design of the student assessment reform

There is increasing awareness among policy makers in OECD countries of reforms or policy initiatives not succeeding or failing on their own merits; their success is dependent upon the process of both the design and their actual implementation (OECD, 2020[8]; May, 2015[51]). This section looks at the design of the Flemish standardised student assessment reform to date based on the information available to the OECD. It discusses the reform's ambitious and innovative design features, starting with the governance of the standardised student assessments, including a discussion on the overarching aim and multiple purposes of the reform. This is followed by an examination of the procedures and methodologies associated with the assessments and concludes with a review of the use of the assessment results.

Governance

Full cohort standardised student assessments to strengthen and monitor the quality of education

Evaluation and assessment provide a basis for monitoring how effectively education is being delivered to students and for assessing the performance of systems, schools, school leaders, teachers and students, among others. Evaluation and assessment in education can serve two primary goals – *accountability* and *development*; both ultimately aim to improve student outcomes. The emphasis placed on each of these goals varies across countries (OECD, 2013[6]). The Flemish school system stands out among OECD countries for its low results-based accountability compared to other OECD countries (see Section 1). Despite this, however, the Flemish government has emphasised the use of the standardised full cohort student assessments as serving the primary goal of development (Flemish Government, 2022[48]; Shewbridge and Köster, 2021[5]).

In recognition of the declining results of the quality of education (Weijts, 2023[52]; OECD, 2019[1]) (see Section 1), the new standardised student assessments aim to strengthen and monitor the quality of education and measure the learning gain at the system-, school-, class- and student levels (Flemish Government, 2023[53]; Flemish Government, 2022[48]; Shewbridge and Köster, 2021[5]). The Flemish student assessments are considered a vital source of information about the quality of education across the Flemish school system; the standardised student assessment data will, for the first time, provide a comparative measure of student performance across all schools in the Flemish Community. More specifically, the student assessments aim to give insight into the extent to which the attainment targets are achieved and to which learning gains occur (Weijts, 2023[52]).

The evidence collected by the OECD confirmed previous findings of general support for the introduction of full cohort standardised student assessments in both primary and secondary education (Shewbridge and Köster, 2021[5]; Molenberghs et al., 2022[54]), with several stakeholders highlighting their potential beneficial use for supporting the professional learning of teachers and school leaders, and school development efforts more generally. The OECD team, for example, had the opportunity to speak with a large group of teacher trainees during an information session on the soon-to-be-introduced standardised

student assessments organised by the DoET (see Annex A). The teacher trainees were generally very positive about the standardised student assessments and how these could help them improve their practice and school development planning more generally.

The benefit of using standardised student assessment results as part of a holistic school quality framework to identify and support underperforming schools

Many education systems, such as those in Ireland, the Netherlands, Singapore and Wales, use a range of data to identify schools on a spectrum of strong to underperforming schools and provide underperforming schools with tailored support to assist them in their school development efforts (Van Twist et al., 2013[55]; OECD, 2018[56]; 2013[6]; Schleicher, 2018[12]). The Flemish Community of Belgium has also adopted such an approach. The full cohort standardised student assessments aim to further aid the Flemish Community in identifying and providing support to underperforming schools.

The government concept note on the standardised student assessments states that:

"*The Flemish assessments are complementary to the foundations of the quality policy, expressed in the so-called quality triangle (attainment targets – Education Inspectorate – pedagogical advisory services). Internal quality assurance at school remains the cornerstone of the quality policy, but will be informed in the future by the test results. The schools can call on their pedagogical advisory services to support them in interpreting their results and developing actions.*" (Flemish Government, 2021, p. 7[57]).

As discussed in Section 1, schools use the OK Framework of quality indicators for their quality assurance processes. The Inspectorate of Education and the PDB also use it – the "quality triangle approach" (see Figure 3).

The government concept note on the standardised student assessment reform also states the intention of ensuring faster – and mandatory – follow up support to those schools whose student performance falls below expectations (Flemish Government, 2021[57]). The full cohort standardised student assessment data will (for the first time) provide a standardised comparative measure of student performance across all schools in the Flemish Community. As such, it will allow identifying underperforming schools and providing them with the necessary support in their developmental efforts, thereby further strengthening the quality triangle approach.

Although the stakeholders interviewed by the OECD were generally supportive of this measure, some raised concerns about the possibility of student assessment data dominating the Inspectorate's school evaluations. The discussions with the Inspectorate confirmed that the OK Framework would continue to be "the" reference guide for all its school inspections and other activities (see Box 2). Inspectors were keen to note that the student assessment data (only) relate to 2 of the 37 indicators of the OK Framework and indirectly relate to an additional 3 indicators regarding quality expectations (see Box 3). Therefore, its influence on the broader interpretation of the quality of education should not be overestimated. This message was reiterated in the draft decree on the standardised student assessment (Flemish Government, 2023[53]) and should be communicated widely, as it will likely help reassure any school leaders, teachers, parents and other stakeholders that may have concerns or misunderstandings.

Box 3. Indicators of the OK Quality Framework that concern student assessment data

Two indicators of the OK Framework directly concern student assessment data:

1. **The school achieves the minimum desired output with the largest possible group of students.** The minimum desired output includes pursuing the development goals and achieving the learning outcomes, basic competencies and curriculum goals. This is related to the pursuit of well-being and involvement of learners and teachers. The general quality culture in the school and the quality of the pedagogical relationship between learners and team members are key elements in this regard.
2. **The school strives for as much learning gain as possible for every student.** The school must make every effort to offer every learner as many development opportunities as possible. It aspires to broad development and strives for maximum learning gain for every learner. This refers to the increase in knowledge and skills and the development of learners' attitudes, competences and talents over a certain period of time.

Several other indicators in the OK Quality Framework relate to student assessment data, i.e. their successful implementation could benefit from using student assessment data. These include:

- **The school evaluates its work cyclically, systematically and reliably based on the results and effects among the learners.** Schools systematically examine, guarantee and improve their quality in education. They specify for themselves the content and the method for their own quality care based on their observations, priorities and goals. The school uses various relevant sources and involves stakeholders when analysing and evaluating the quality of its education. The results and effects that the school achieves with the learners form the starting point in this regard, together with the context and input characteristics. We can only speak of "systematic" work if the school evaluates its work regularly. In addition to the systematic nature, the cyclical nature of quality care is important. A cycle does not just repeat what has happened before, but adds something new and novel. In this way, quality is developed further with each cycle.
- **The school develops and pursues an effective policy in teaching and learning.** The school supports and monitors the continuity, construction and coherence of the objectives; the design of the teaching process; and the living and learning environment, guidance, and evaluation and reporting practices. It collects targeted information to secure and adjust its work.
- **The school team evaluates the teaching/learning process and the achievement of the goals in a broad and informed manner.** Evaluation is an essential and integral part of the learner's learning process. The school team adopts an informed view of evaluation. Evaluation is broad and representative of the learner's own goals and the validated target framework. The school team carefully chooses various evaluation procedures and instruments in order to collect reliable information. The evaluation criteria are communicated objectively and clearly. Evaluation is stimulating and development-driven, and is conducted fairly and transparently.

Source: OK Education Quality (2016[24]), *The reference framework for Quality in Education: quality expectations and quality images,* https://www.onderwijsinspectie.be/sites/default/files/2022-06/OK_magazine_eng.pdf.

Measures to prevent school ranking while ensuring access among key "users"

A concern expressed by some of the Flemish stakeholders was that the student assessment results might be used to create school rankings or league tables (Penninckx, 2019[58]; Delegates of the Education Providers, 2022[59]). Research evidence shows that although the publication of actual student assessment results provides important information, they – on their own – are poor measures of school performance (OECD, 2008[60]). For example, they fail to take into account students' prior achievement levels and the broader school context (OECD, 2013[6]). Therefore, in response and following the examples of education systems such as the French Community of Belgium and the Netherlands, the Flemish government has taken legal measures to ensure that the results of a school are not made public (i.e. the standardised student assessments do not fall under the open government policy) and prevent school rankings from being drawn up.

The draft decree on the standardised student assessments states that schools themselves will have access to the assessment results. Their school feedback reports will contain contextualised data, taking into account the characteristics of the school's student population. In this way, schools should receive detailed data and information that can be used for quality assurance and development purposes. The school-level results will be shared with the education providers and the Inspectorate of Education, so that they have an additional source of information. The Inspectorate can carry out an inspection visit on the basis of repeated poor performance on the Flemish tests or a decrease in learning gains.

A school can provide access to its school feedback report to parents of students who took the test. The school board determines the procedure and modalities for this. The parents, in turn, are not allowed to communicate the results to third parties. If they violate this duty of confidentiality, they can be sanctioned with a fine (Weijts, 2023[52]).

Prioritisation of purposes to support the test design process

As mentioned above, the overarching aim of standardised student assessments is to strengthen and monitor the quality of education by achieving attainment targets and measuring the learning gain at multiple levels. The student assessment results are to be used for multiple purposes:

- At the system level, monitoring the extent to which the attainment targets are achieved, to which learning gains are generated and as an element of quality assurance at the system level.
- At the school level, as an element of internal and external quality assurance:
    - As one of the elements in the internal quality assurance of the school. To this end, the school and the school board will be able to consult their own feedback report with contextualised results in a secure manner.
    - As one of the elements in the functioning of the PDB.
    - As one of the elements in the functioning of the Inspectorate of Education.
- At the level of the student group or class, as one of the elements for reflection on the pedagogical-didactic action of the school team.
- At the student level, as one of the possible results that the class council can take into account in the student's final marks. The standardised student assessment results are, as such, not used as the sole criterion for students' final marks for that subject (Flemish Government, 2023[53]).

Across OECD countries, it is common for standardised student assessments to serve multiple purposes (OECD, 2013[6]). This is known as "purpose pluralism" (Newton, 2017[61]). However, critics have warned of the risks of purpose pluralism (Penninckx et al., 2017[62]; Morris, 2011[63]; Eurydice, 2009[64]). As noted in the Standards for Educational and Psychological Testing,[1] "*Most educational tests will serve one purpose better than others; and the more purposes an educational test is purported to serve, the less likely it is to serve any of these purposes effectively"* (AERA, APA and NCME, 2014, p. 188[65]). Since different purposes often require different test design decisions, purpose pluralism needs to be managed (Newton, 2017[61]; Hemker and Feskens, 2021[66]). Policy makers and assessment developers need to be clear from the beginning about the primary purpose for which assessment results will be used. It is essential to prioritise explicitly among purposes. The primary purpose and, to the extent possible and desirable, other high priority purposes, should then determine assessment design characteristics to ensure the validity of results for intended purposes (OECD, 2013[6]). Without such prioritisation, the test design is at risk of being suboptimal (Somers, 2021[67]). For example, a test primarily designed to monitor system-level changes in

[1] The 1966 published Standards for Educational and Psychological Testing is often considered the gold standard in guidance on testing in the United States and in many other countries.

performance does not need to be as reliable at an individual student's level as a test primarily designed to measure performance at the student level. One of the key drivers of the reliability of a test is its length (Revelle and Condon, 2019[68]). If the test is primarily designed to monitor system-level changes, it can be much shorter than a test designed to identify strengths and challenges per learning area at a student level. This point also applies whether the student- or system-level focus is on raw outcomes or value-added/learning gain outcomes.

Another example of a design feature that could be optimised through purpose prioritisation is content coverage, i.e. of topics and/or competencies included in the assessments. If the assessment is primarily designed to monitor system-level changes, then each individual test can merely cover a sample of the topics and/or competencies of the subject assessed. For example, while some students could be assessed on geometry and fractions on a mathematics assessment, others could be tested in statistics and arithmetic. Giving students such different test versions can (still) ensure that the entire subject content is tested at the system level and provides statistically valid results.

If, however, the primary purpose of the assessment is to provide classroom or individual student-level feedback, then a more comprehensive sampling of the subject content is preferable; for example, by including all (key) topics and/or competences of the subject assessed in all tests. This broad(er) coverage of subject content would make the assessment outcomes more valuable to teachers and students. Students would be better able to understand their strengths and weaknesses, as they would have been assessed in all (key) topics and/or competences. Teachers could also better compare outcomes across different students. The issue of content sampling is discussed further below.

A final example relates to the publishing of assessment results and the release of tests at the end of each assessment cycle. If the primary purpose of the assessment is to monitor system-level performance, then the maintenance of test standards, test security and confidentiality are paramount. Although it is possible to release tests (as done in England, for example), it significantly complicates the test development process – involving a cycle of pre-testing items and/or using anchor items. If, however, the primary purpose of the test is to provide feedback to students and teachers, then the release of the tests might take precedence and a more resource-intensive and complex approach to test development and the maintenance of standards should be considered.

To conclude, without prioritisation of purpose(s), it is possible that the design of the assessments will not meet expectations for their intended use. There is, therefore, believes there is a need to clarify the main purpose or purposes of the standardised student assessment, as this would guide the best use of alternative and/or complementary testing strategies for achieving the desired purpose(s). Prioritisation will be further elaborated below.

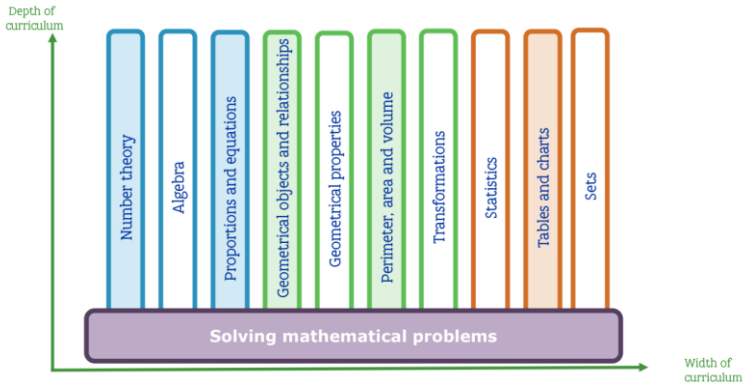
Ensuring strategic oversight of assessment design and administration decisions

The Flemish government commissioned the Steunpunt to support the introduction of the standardised student assessment for the period 2021-25. A steering group has been established to oversee and guide its work. The design process of test items for mathematics and Dutch language reading comprehension and grammar has been shaped through a comprehensive process of stakeholder engagement and co-construction. This can be considered a clear strength of the reform approach elaborated on in Section 3.

While much of the Steunpunt's work is detailed and highly technical, the decisions of the Steunpunt and its steering group have significant implications for policy and practice. The OECD identified a number of test design and administration decisions that will affect how the results can be used. First, the mathematics assessment design suggests that an implicit prioritisation of the purposes of the assessments has taken place for this subject, even if this has not yet been communicated. On the contrary, decisions on the Dutch language test design may suggest a different prioritisation.

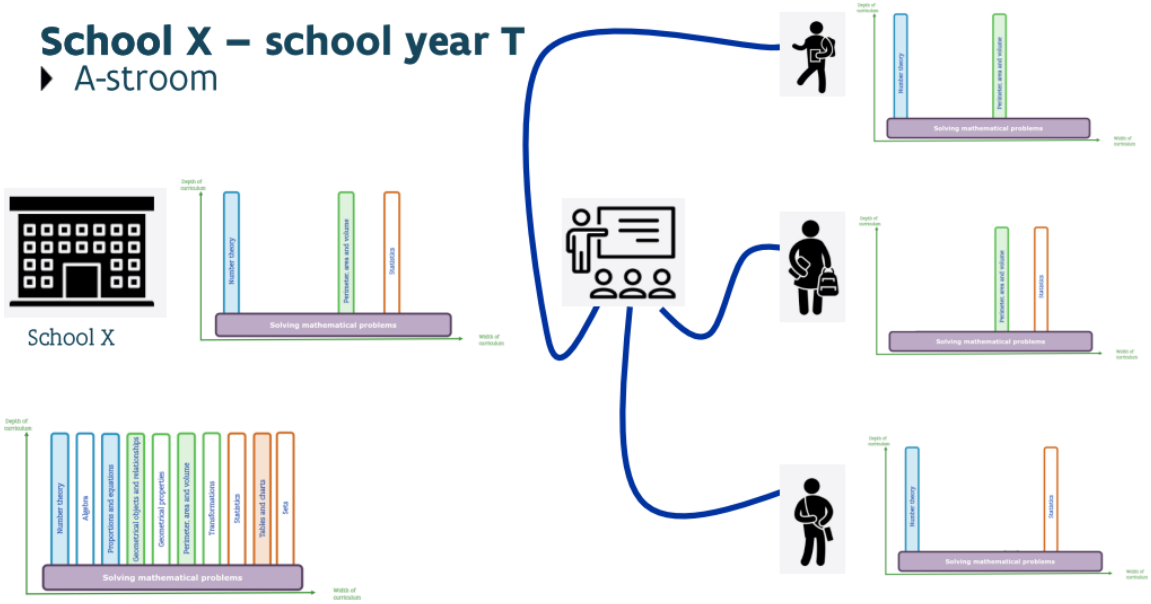
As a starting point, the decision was taken to assess ten key secondary education curriculum competences for the A stream mathematics assessment (i.e. for the general education stream in secondary education) (see Figure 7). The computer-based administration of the assessment allows to randomly select three out of the ten competencies that are assigned to a school, for example, “number theory”, “geometry” and “statistics” (Figure 8). From these three competences, each student from that school will be tested in two randomly assigned competences, for example, “number theory” and “statistics”. As a result of this approach, the curriculum competencies assessed will vary among students in the Flemish system and even within schools. In this example, two-thirds of the students in that particular school will get tested on the competence “number theory”, two-thirds of the students will get tested on the competence “geometry” and two-thirds of the students will get tested on the competence in “statistics”. Other schools could be assigned three different mathematics competences out of the ten available.

Figure 7. Competencies tested in the A stream secondary mathematics assessment



Source: Presentation provided by the Steunpunt during OECD team visit to the Flemish Community in June 2022.

Figure 8. Overview of the secondary education mathematics assessment design



Source: Presentation provided by the Steunpunt during OECD team visit to the Flemish Community in June 2022.

The Dutch language assessment has a different design, however. Unlike in the mathematics assessment, all students will be assessed on the same competencies in the Dutch reading comprehension and writing assessments, which include testing students' grammar. The reading comprehension assessment will consist of different language tasks in which students will get a few texts to read with multiple-choice questions. In both, the Dutch reading comprehension and the mathematics assessments, students may be asked different questions, as they will vary on difficulty depending on the student's performance on the previous questions of the assessment (i.e. they are adaptive assessments). In the Dutch writing (and grammar) assessment, all students will also be assessed on the same competencies (contrary to the mathematics assessment), and it will not be adaptive; therefore, all students will be asked the same questions.

Apart from the challenge of communicating the difference between these test designs to schools, students and their parents, the test design features also determine the feedback that can be provided and the making of certain comparisons based on the mathematics assessment. For the student results on the mathematics assessment, at the class level, the teacher will receive information on how the class is doing on (only) three competences. This design feature limits the use of the data for making comparisons and reflections by teacher teams, especially in smaller classes. The fact that the assessed competences differ among students also hinders their use for student marking as part of students' final marks.

At the school level, the school receives annual feedback on three mathematics competences. These are intended to partially overlap with those assessed in adjacent years. This approach aims to balance providing the school with feedback on a broad range of competences and giving the school a tool to monitor trends over time in support of its internal quality assurance. Although such an approach further limits the risk of using these data for school rankings (which will be forbidden by law), it also affects the use of student assessment data for making school-level comparisons. It constrains the use of these data for internal and external quality assurance and helps identify and support those schools in need of support – three of the purposes of the standardised student assessment reform. The design decisions taken for the mathematics assessment as such suggest an (explicit or implicit) prioritisation of the assessment purposes for the use of the assessment results for system-level monitoring, or possibly a less systematic approach has taken place.

Although these choices are not indefensible or wrong in any way, they have profound consequences for using the assessment results. Therefore, all stakeholders must understand these consequences and be aware of their strengths and limitations (Somers, 2021[67]). It would be helpful to clarify and communicate the rationale for these decisions and explain how they align with and support the realisation of the main purpose(s) of the Flemish standardised assessments. This could increase support among stakeholders.

Second, the decision to not release the test items and questions to schools after the students have completed the student assessments limits the feedback that can be provided to teachers and students. After a discussion with representatives of the Steunpunt, the OECD learnt the decision was taken to allow for the reuse of all the test items in the following years. As mentioned earlier, this practice allows test standards to be more easily maintained and for more accurate monitoring of trends over time. On the other hand, it limits the feedback to support teachers in their reflections on their teaching and students' learning. Even if feedback is provided on each of the tested competencies, teachers would still benefit from being able to explore in detail which type of items students succeed or struggle with.

In light also of the low data literacy skills of teachers and the identified challenges in their ability to collaborate to make the most of the available assessment data to improve their teaching (Shewbridge and Köster, 2021[5]; OECD, 2021[4]), the DoET and/or the Steering Group may consider revisiting this decision. They may look towards the examples of Denmark and British Columbia (Canada), which share the tests together with feedback reports to support reflection on pedagogical didactics by the school teams and to inform internal quality assurance processes (i.e. support the realisation of two of the reform objectives). In

Denmark, for example, the databases with assessment results contain all items used for each student and the answers they gave, which school professionals can access for pedagogical purposes (Wandall, 2017[69]). Also, in British Columbia (Canada), the reports of the provincial Foundation Skills Assessment for the student and teachers (which can also be shared with parents) include the whole test booklet with responses and student performance on proficiency levels with descriptive, strength-based language (Government of British Columbia, n.d.[70]).

As noted in the feasibility study on the Flemish standardised student assessments, feedback reports are arguably one of the most important elements of student assessments for supporting improvement in teaching and student learning (Somers, 2021[67]). If the Flemish Community of Belgium decides not to release the test questions, it should ensure schools and teachers still receive adequate and accessible feedback and support for interpreting the student assessment results (see below) (Welsh Government, 2021[71]).

As mentioned above, these examples are by no means a criticism of the work undertaken by the Steunpunt. The point is simply to show the implications of technical design decisions for the actual implementation of the reform. That said, there is a need to clarify the main purpose and possible other high-priority purposes of the standardised student assessment. As discussed above, the test design decisions taken for the mathematics assessment suggest an (explicit or implicit) prioritisation over the uses of the assessment results, or possibly a less systematic approach has taken place. This is an issue for all partners involved (the Steunpunt, the Steering Group, the ministry and the Ministerial Cabinet) to clarify. If indeed a prioritisation of purposes has taken place, this should be communicated widely.

On the other hand, if no prioritisation has taken place, the DoET should conduct a prioritisation exercise of the assessment purposes with all stakeholders to inform the ongoing and future design efforts and uses of the assessment results.

It should communicate the main purpose(s) of the standardised student assessment to ensure all stakeholders have a broad understanding of the assessment and of the need to combine a variety of assessment information to come to a good judgement and valid interpretation of student learning and progress. For this, it may look towards the example of Canada, which in 1993 released the influential *Principles for Fair Student Assessment Practices for Education in Canada*. This document outlined the key elements for assessment practice and since then has served as the foundation for teacher handbooks, board policies and departments of education policy documents on assessment and test development in all Canadian jurisdictions. The principles and guidelines were intended for both assessment practitioners and policy makers to identify the issues to be taken into account so that assessment exercises can be deemed fair and equitable. The text has acted both as a set of parameters and a handbook for assessment (OECD, 2013[6]).

Furthermore, given the importance of the implications of technical decisions and to avoid unintended consequences, there seems to be a need to clarify the roles and responsibilities of the Steering Group and to strengthen oversight by directly involving the ministry and Ministerial Cabinet in decisions on key issues. For issues of particular importance, like the proposed prioritisation of purposes, the Steunpunt could suggest different scenarios, explaining their respective strengths, limitations and implications, for the Steering Group and (then) the ministry and Ministerial Cabinet to decide on.

Procedures

Standardised assessments focused on Dutch language and mathematics

As mentioned above, the Flemish standardised student assessments will focus on assessing students in Dutch language (i.e. reading comprehension, writing and grammar) and mathematics, responding to the

decline of results in literacy and numeracy in international and national student assessments. The Flemish government recognises this is only a limited part of the school curriculum, but in the short term, considers that it is not organisationally feasible to further expand the Flemish student assessments.

Considering the limited content focus (i.e. on Dutch language and mathematics), the DoET may want to monitor for unintended consequences, such as narrowing of the curriculum or schools allocating more resources to these subjects that are tested (Copp, 2018[72]; Ro, 2019[73]; Darling-Hammond, 2020[74]; OECD, 2013[6]). Although the standardised student assessments are intended to be "low stakes", evidence shows that it is impossible to fully control the stakes associated with any student assessment. Whatever the intention regarding the stakes of standardised student assessments, negative effects can grow over time, as was, for example, the case for national assessments in primary schools in England (Perryman et al., 2011[75]; Wyse and Torrance, 2009[76]) and the National Student Achievement Tests in Lithuania (Raudienė, Kaminskienė and Cheng, 2022[77]). Small changes in the use of assessment data or even how assessment results are communicated can impact on the perceived stakes of student assessments (Feniger, Israeli and Yehuda, 2015[78]). This section will elaborate on both issues (i.e. the use of test results and communication).

However, once the Flemish standardised student assessments are well-established, the Flemish government should (as intended) examine the option of expanding the assessment to other subjects, learning areas or competences. This would provide further comparative student assessment data for monitoring progress against the breadth of the school curriculum (i.e. attainment targets) and further support developmental efforts at different levels of the system. This could also diminish the risk of unintended consequences, such as the narrowing of the curriculum.

Defining how learning gains will be measured and how contextual factors will be taken into account

Similar to OECD countries such as Australia, England (United Kingdom), Italy, the Netherlands and New Zealand (OECD, 2013[6]; Leckie and Goldstein, 2017[79]; ACARA, 2021[80]), the Flemish Community intends to use the standardised assessment results to measure learning gains. It aims to measure learning gains at system-, school-, and student level (Flemish Government, 2022[48]; 2021[57]). According to the Steunpunt's current definition, learning gain refers to a student's learning progress across years. At a school level, this will be calculated as the average learning gain of the students in the school. As the measure is longitudinal (i.e. same student/school across time), this estimate can only be calculated for the first time in 2026, with the cohort of students who will take the test in Grade 4 in 2024 and again in Grade 6 in 2026. Given the complexity of the student assessment reform and the necessary management of expectations, there seems to be much to gain from clarifying and communicating the possibilities and limitations of using the data for measuring learning gains, especially in the initial years.

Furthermore, in their discussions with the OECD, some stakeholders commented that there is yet to be a clear explanation of how "value-added" results will be defined and measured. There also seems to lack clarity among some stakeholders regarding the difference between value-added and learning gains; an issue also raised by the Flemish Education Council (2022[81]). This lack of clarity risks undermining trust in the use of the assessment data, not least because decisions about the calculation of value-added measures can be controversial (Brown, McNamara and O'Hara, 2016[82]; Munoz-Chereau, Anwandter and Thomas, 2019[83]; OECD, 2008[60]; Schiltz et al., 2018[84]; Levy et al., 2019[85]; Amrein-Beardsley and Holloway, 2017[86]), as will be elaborated upon below.

School attainment measures aim to report students' average performance at the end of the academic year or an educational stage. These measures do not accurately reflect the quality of schools, as they confound quality with the composition of each school's intake. Schools with higher attaining students at intake tend to also score higher in measures taken at the end of schooling, irrespective of the effectiveness of the

schooling provided. Value-added measures are widely considered a fairer and more meaningful way to compare the effectiveness of schools, as they implicitly attempt to adjust for what can be substantial differences in the composition of students' prior attainments and other characteristics between schools at intake (Leckie and Goldstein, 2017[79]). In the interviews with the OECD, some stakeholders noted the different contexts and provisions for schools in terms of the composition of the student population. When stakeholders referred to differing student population composition across schools, they pointed out the need for adequate contextual understanding to interpret results on educational outcomes meaningfully. The use of value-added measures in the Flemish Community is, therefore, broadly welcomed by stakeholders.

There are, however, many ways in which value-added measures can be calculated, and in some countries, these measurements have evolved over time. For example, in England, the headline measure of school progress has changed many times: "value-added" (2002-05) to "contextual value-added" (2006-10) to "expected progress" (2011-15) to "Progress 8" (2016-) (see Box 4). In addition, the different ways in which value-added measures are calculated vary in reliability. Whichever method is chosen, it would be important to report the statistical uncertainty inherent to the measures (Leckie and Goldstein, 2019[87]). They also vary in their validity and require assumptions to be made which are both technical and political in nature. For instance, adjusting for socio-economic and other demographic variables of students when calculating school-level results can be done accurately with indicators and could be regarded as a "fair" measure. At the same time, it implicitly implies a lower expectation of students with certain characteristics or context (e.g. lower socio-economic background), which may not be acceptable in certain countries (e.g. England). The political and social dimensions and implications should be reflected upon and not be overlooked, as should the consequences of these decisions. For this reason, choices about the design of the value-added measures would ideally be made in consultation with not only statistical experts, but a wide range of stakeholders, including school leaders and teachers. This will enable a thorough understanding of the consequences of the design of the measures and will also facilitate their acceptance in the educational field and beyond.

Box 4. The use of value-added measures of school effectiveness in England and Italy

England

In England, the simple value-added measure used from 2002 to 2005 was criticised for failing to account for school differences in student socio-economic and demographic characteristics, which had been shown to predict outcomes at the end of schooling even after adjusting for prior attainment at the end of primary education. The measure was therefore thought to be biased in favour of schools with more socially advantaged intakes.

The contextual value-added measure, used from 2006 to 2010, attempted to better separate school effectiveness from the composition of their intakes. The measure took into account the prior attainment of students but also their age, gender, ethnicity, socio-economic status (as proxied by free school meal eligibility) and various other student and school characteristics. In 2011, this measure was abandoned, ostensibly because it was difficult for the public to understand and by adjusting for school differences in students' socio-economic and demographic backgrounds, the measure entrenched low educational aspirations in disadvantaged pupil groups (UK Department for Education, 2010[88]).

In its place, two new school progress measures were introduced and used from 2011 to 2015. The first was a value-added measure based on predictions of students' outcomes at age 16 based on their prior attainment at the end of primary school. The second was a measure of expected progress. The measure, which was reported separately for English language and mathematics, was calculated as the percentage of students in each school who "make the progress expected of them" during secondary schooling, defined for all students as three (or more) national curriculum levels. Thus, for example, students achieving Level 4 in English at the end of primary school were expected to progress three national curriculum levels in that subject by age 16. The measure did not take into account school differences in students' socio-economic and demographic backgrounds.

This measure was replaced from 2016 with the Progress 8 measure. Progress 8 is defined as a student's total point score measured across the General Certificate Secondary Education (GCSE) English and mathematics and six further subjects (from an approved list). Schools' Progress 8 scores are simple averages of the differences between students' scores on the eight GCSEs and the national average scores of students on the eight GCSEs with the same prior attainment at the end of primary schooling. There are no adjustments for students' socio-economic or demographic characteristics. Each measure has its benefits and disadvantages and should be interpreted with caution, given the statistical issues of quantitative school comparisons. It is important to note that schools whose intakes are more socio-economically advantaged are more likely to benefit from Progress 8 than from others used in the past (Leckie and Goldstein, 2017[79]).

Italy

In Italy, national standardised assessments take place annually in primary and secondary schools to assess student performance in mathematics, English and Italian, as well as Slovenian in a few small territories. The National Institute for the Evaluation of the Educational System of Education and Training (INVALSI) is responsible for the design and data processing of the assessments. Results are provided at a class, school, regional and national level.

There was strong criticism from teachers when the assessments were first introduced, as test results were not considering students' background, which could be a highly influential factor in their performance. As a result, INVALSI introduced new statistical tools to calculate a school value-added estimation measure. This takes several student-level characteristics into account, including gender, variables related to the personal and socio-economic background (e.g. index of the social, economic and cultural status of the student's family, etc.) and others related to the student's scholastic profile (e.g. prior score at INVALSI test, Grade repetition, etc.). School-level variables related to the school context, such as the number of enrolled students, are also included in the estimation. To facilitate the applicability of results, INVALSI uses the indicator to create categories which are easily interpreted by schools (see Table 3). The categories use the observed scores and calculated school value-added estimate to indicate the school's performance (i.e. its effectiveness) with respect to the national level or the reference region.

Table 3. INVALSI table with school performance categories based on observed and school value-added performance

	Positive school effect	Slightly positive school effect	School effect equal to the average	Slightly negative school effect	Negative school effect
Observed score above the average	School contribution very evident Results good	School contribution evident Results good	School contribution in the average Results good	School contribution not adequate Results good	School contribution inadequate Results good
Observed score equal to the average	School contribution very evident Results acceptable	School contribution evident Results acceptable	School contribution in the average Results acceptable	School contribution not adequate Results acceptable	School contribution inadequate Results acceptable
Observed score below the average	School contribution very evident Results to be improved	School contribution evident Results to be improved	School contribution in the average Results to be improved	School contribution not adequate Results to be improved	School contribution inadequate Results to be improved

Sources: Leckie and Goldstein (2017[79]), The evolution of school league tables in England 1992-2016: 'Contextual value-added', 'expected progress' and 'progress 8'; Cardone, Falzetti and Sacco (2019[89])*Working Paper N. 43/2019: INVALSI data for school system improvement: the value added;* INVALSI (n.d.[90]), *The INVALSI tests according to INVALSI,* https://invalsi-areaprove.cineca.it/docs/2018/INVALSI_tests_according_to_INVALSI.pdf.

The Steunpunt plans to adjust for students' backgrounds, but the variables included in the indicator have yet to be decided. Therefore, the Steunpunt should, as planned, engage with a broad range of stakeholders to develop the Flemish definition and operationalisation of valued-added student performance. It may look towards the examples of other countries (see Box 4) and the earlier undertaken feasibility study on the Flemish standardised student assessments (Somers, 2021[67]) to identify the strengths and limitations of different definitions and measurement strategies and work towards valued-added measures that fit the

Flemish context. The key partners involved, i.e. the Steering Group, the ministry and the Ministerial Cabinet, should agree upon the proposed definition of value-added student performance to ensure that all political and social dimensions are considered.

Ensuring continuous alignment between standard student assessments and attainment targets

Improvements in educational quality are most strongly driven when there is strong alignment between educational goals, learning targets, and assessment content and standards (Shannon and Bylsma, 2007[91]; Looney, 2011[92]; Care et al., 2018[93]). A critical element in the effectiveness of standardised assessments in improving student learning is their alignment with educational goals and attainment targets (OECD, 2013[6]; Biggs, 1996[94]). When this is the case, teaching and learning are clearly focused on agreed objectives, which are then assessed. This requires that any changes to attainment targets be duly reflected in assessment content and that teachers and students clearly understand the education goals and attainment targets. As the usefulness of standardised assessments in improving student performance can be undermined by misalignment between assessment content and attainment targets, the design of the assessments may need to be revisited periodically to ensure they are sufficiently aligned. This is because assessments are known to influence both teaching and learning – i.e. they have a "washback effect" (Alderson and Wall, 1993[95]; Cizek, 2005[96]). Washback can have positive effects on learning when assessments are aligned with learning goals but can be negative when alignment is lacking.

As mentioned above, in June 2022, the Belgium Constitutional Court decided that the new secondary attainment targets violated the fundamental right to freedom of education and were therefore annulled (Constitutional Court, 2022[47]). The new secondary attainment targets had been introduced for the school year 2021/22 in phase 1 already. The Court, therefore, provided a transitional phase in which schools may continue to use new secondary attainment targets until the 2024/25 school year, allowing time to amend the targets.

The DoET should ensure continuous alignment between the student assessment content and the attainment targets over time. The quest for continuous alignment is common across education systems. New South Wales (Australia), for example, has recently embarked on a reform of its school curriculum and is currently examining the impact of the new syllabi and pedagogical approaches on its assessment and evaluation arrangements, including the use of the National Assessment Program – Literacy and Numeracy (NAPLAN) student assessment results (Masters, 2019[97]).

An innovative reform that would benefit from a "learning-by-doing approach"

The intended design and delivery of the full cohort standardised student assessment is ambitious and innovative, among others, because of the computer-based and adaptive nature of the assessments. There is value in evolving standardised student assessments over time based on lessons learnt from their implementation (Shewbridge and Köster, 2021[5]). Such a "learning-by-doing approach" may benefit the successful implementation of the student assessments, as elaborated below.

Computer-based student assessments

The Flemish standardised student assessments have been developed as computer-based assessments, in line with developments in other OECD countries like Australia, Denmark, England (United Kingdom), Finland, Israel, Latvia and New Zealand that have and/or are in the process of switching to computer-based student assessments. Computer-based student assessments have advantages and disadvantages. Some of the most obvious advantages include the possibility of rapidly marking them and providing feedback, reducing human biases and errors, and the fact that they, at least partially, address teachers' workload (Leaton Gray and Kucirkova, 2021[98]). Other more technical advantages include the possibility of using

adaptive testing or using sophisticated trackers of academic progress across an extended period of time (Veldkamp and Sluijter, 2019[99]).

On the other hand, there are challenges to consider, such as how data should be collected, used and stored (Timmis et al., 2015[100]). Ethical questions may also emerge in relation to inequalities, as it cannot be assumed that all teachers and students have the necessary skills to use technological devices/software. This may impact on students' assessment results (Wyatt-Smith, Lingard and Heck, 2021[101]). Other challenges arise with specific technology-enhanced assessments. For example, assessments marked by algorithms may be less transparent and hard to verify, particularly with machine learning (e.g. essay marking) (Veldkamp and Sluijter, 2019[99]).

In its review of the challenges of computer-based testing in England, Ofqual studied its introduction in Finland, Israel and New Zealand (Ofqual, 2020[102]). The authors found that successful implementation required a strong "risk appetite" – recognition and acceptance by stakeholders that some things would go wrong and an understanding there would be some "learning by doing". Each of the countries took a different approach to implementation. For example, New Zealand took a gradual, voluntary approach. This resulted in a longer timeline for introduction but enabled public perception to change as the adoption grew to ensure positive user perception prior to use. Technical glitches were easily overcome by the ability to revert to paper-based testing. The downside of this approach was the dual running of paper and computer-based assessments, which risks an unfair advantage or disadvantage to groups of students. Biases related to the method of assessment delivery could ultimately hinder performance comparisons (McClelland and Cuevas, 2020[103]; Kolen and Brennan, 2014[104]; Sandene et al., 2005[105]). Another example is provided by Australia, which underwent a broad research and development programme before and during the transition of its NAPLAN to a computer-based delivery, providing useful results that may inform the Flemish student assessment reform (see Box 5).

The Flemish Community opted for only offering computer-based student assessments, which avoids a bias related to the assessment method. Evidently, computer-based assessments need technical preconditions, which include access to enough devices of a consistent specification; sufficiently reliable Internet and/or local network capabilities; sufficient staff with the expertise to support the introduction and ongoing use of computer-based testing; physical spaces with the electrical and network facilities suitable for numbers of students to take the assessments on devices concurrently; and strong IT infrastructure to manage security risks. An offline modality of taking the test may provide a means for minimising some of these risks, but this requires some additional time and guidance to schools for installing the software beforehand and limits the use of tools and the adaptability of assessments (Somers, 2021[67]).

It is possible to meet these requirements, though they require an initial financial investment and support for ongoing maintenance. The DoET, with the support of European Structural Funds, is dedicating EUR 385 million to a digitalisation project across the education sector (Government of Flanders, n.d.[106]; SGI Europe, 2021[107]). This includes investment in infrastructure (e.g. laptops, computers, software packages), Internet connectivity and teacher training. This project ("Digisprong") is currently on schedule, with completion expected in 2024 (Goverment of Flanders, 2022[108]).

Box 5. Research findings guiding the transition of National Assessment Program – Literacy and Numeracy to an online format in Australia

The Australian Department of Education and Training funded a series of studies to ensure technical and content readiness for the transition of the annual National Assessment Program – Literacy and Numeracy (NAPLAN) to an online assessment. These studies were conducted by the Australian Curriculum, Assessment and Reporting Authority (ACARA), an independent statutory authority. They covered different aspects of the tests, including: trialling the new test designs, the impact of an online delivery mode, students' engagement, etc. The research agenda included diverse study designs, such as literature reviews, trials in diverse schools and cognitive interviews. Some of the findings are discussed below:

- **Device suitability:** The assessments could be taken on laptops and tablets without overall device effects impacting performance and without the need for an external keyboard. In this regard, students' familiarity with the device was a key factor.
  ACARA also performed piloting exercises to evaluate technical elements such as item display, navigation options and screen displays, whose findings were incorporated into the design.
- **Tailored test development studies:** Studies on the design of multi-stage branching tests (i.e. adaptive tests depending on students' performance on previous items) indicated that this type of design offered precise measurement, particularly for students scoring on the high or low end of the continuum. The design allowed for better catering to students' assessment and learning needs. Psychometric factors such as the test's boundaries (i.e. performance cut-off point for easier or harder branching items) needed further refining.
- **Automated essay scoring:** Empirical evidence supported the feasibility and validity of the use of automated essay scoring, as well as indicating no operational barriers to its implementation. All written samples included were in English, as this was the language assessed.
- **Student engagement:** Students in the trials found the assessments more engaging than the previous paper-based tests. No students felt penalised for typing their responses and perceived the assessment as fair and easy to navigate.
  Regarding the adaptive nature of the assessments, cognitive interviews showed that students were not distracted by the increasing difficulty of test items. One study also highlighted that this design enabled educationally disadvantaged students to be engaged and to generate a sense of accomplishment. The most challenging items were perceived as appropriate for students with higher levels of achievement. Overall, students completing the adaptive tests reported a more positive subjective experience than those who were given a fixed exam.
  These findings shed light on the potential benefits for students' engagement and motivation, and for the precise measurement of students' performance of these innovative methods. They also highlight the feasibility of computer-based assessments in schools for students as young as Year 3 (i.e. normally students aged 7-8).

Sources: ACARA (2014[109]),*Tailored test design study 2013: Summary research* report, https://nap.edu.au/docs/default-source/default-document-library/tailored_test_design_study_2013_summary_research_report.pdf; ACARA (2015[110]), *NAPLAN Online Research and Development. Report 1: Device Effect Study – Literature Review. Report 2: Device Effect Study – Field Trial*, https://nap.edu.au/docs/default-source/default-document-library/naplan-online-device-effect-study.pdf, Lifelong Achievement Group & Martin, A.J., (2015[111]), *Online NAPLAN Testing and Student Motivation: Exploring Adaptive and Fixed Test Formats*, https://www.nap.edu.au/_resources/Online_NAPLAN_and_Student_Motivation.PDF, ACARA (n.d.[112]), *NAPLAN: Understanding online assessment. Research and development*, https://www.nap.edu.au/naplan/understanding-online-assessment/research-and-development#SEM, Lazendic, Justus and Rabinowitz (2018[113]), NAPLAN online automated scoring research program: Research report, https://nap.edu.au/docs/default-source/default-document-library/naplan-online-aes-research-report-final.pdf, all accessed on the 20 September 2022.

Various stakeholders reported that not all IT challenges may have been resolved in schools across the Flemish Community. There were concerns that some of the purchased devices may not always optimally work with the test software. The DoET, however, ensured the OECD that every device will be compatible with the testing software. This is essential to prevent unequal performance of different devices during the assessments, which risks creating unfairness for students without access to the most up-to-date technology and hindering the comparability of results. It will also be vital to prevent unequal opportunities for students to practise on the relevant software or devices beforehand. The 2022/23 school year could be used to ensure adequate IT infrastructure is in place. The large-scale piloting will also provide valuable insight into the challenges schools face when administering the assessments, as well as showcasing what went well – both providing valuable lessons for their administration in following years. If the piloting shows

that IT challenges continue to be an issue for (many) schools, a fall-back option could be, in the first year(s), to allow these schools to administer the tests in paper-based format, with all other schools administering the tests through a computer-based format. Of course, paper-based tests are not adaptive.

Furthermore, it will be important to communicate beforehand that the administration initially may not go perfectly, noting that this is to be expected. Lessons learnt from this "learning experience" will be taken forward in the following rounds of student assessment; pointing to the process evaluations that are planned for the piloting and first year(s) of implementation (Flemish Government, 2022[48]).

In interviews with the OECD, some stakeholders raised concerns of the young age of Grade 4 students taking the test. Not all younger students may have the experience, skills and/or confidence to take the tests through a computer-based format. Again, the large-scale piloting will provide the opportunity to explore whether this is the case. Giving all students sufficient opportunity to practise completing tests on the software will be an important part of ensuring the tests are accessible and fair for all.

Adaptive student assessments

Another innovative feature of the assessments is that they are not only computer-based but also adaptive for the mathematics and Dutch language reading comprehension assessments (see Section 1). As mentioned above, the Dutch language writing assessment is not adaptive, however. Computerised adaptive assessments are a form of computer-based testing in which the level of difficulty of the assessment questions adapts to the student's ability. This is done by automatically selecting question items based on the student's response to a previous (set of) question(s). Students who answer questions correctly are directed to a more difficult question and those responding incorrectly receive an easier question (National Council on Measurement in Education, 2019[49]; OECD, 2013[6]). Testing in reading comprehension is to be automatically marked and adaptive in nature starting in May 2024. Dutch writing is planned to be marked through machine learning, though only from May 2025 onwards, at the earliest.

Adaptive testing is efficient, as students do not waste time and effort on questions that are too hard or too easy for them. As candidates only answer items that are paired to their ability, test length can be reduced in comparison to linear test forms, and test administration can, if desired, be more flexible in terms of time as a result of individualised testing (Veldkamp and Sluijter, 2019[99]). Adaptive testing maximises the measurement information gathered. While a standardised test can only provide a snapshot of student achievement in selected targets and subjects, within these limitations, adaptive assessments are able to provide an accurate measure of student performance within a discrete area. As each student sits a different test, including questions adapted to his/her ability level, this can allow for more thorough diagnostic feedback (OECD, 2013[6]). This may also make for a better test experience, as students are not anxious about questions they cannot answer or bored by questions they find trivial.

At the same time, some international evidence shows that test-takers report feeling discouraged after taking a computer-based adaptive test, and this feeling can reduce learning self-efficacy and motivation (Kimaru, 2017[114]). In addition, adaptive assessments require a large bank of test questions, which increases the development costs (OECD, 2013[6]). Furthermore, it may be challenging to explain to students and their parents why test experiences differ between students, as well as how the test results should be interpreted (Somers, 2021[67]).

Given that the large-scale piloting in May 2023 will not have the adaptative testing feature, the DoET should – as planned – use the first administration of the adaptive assessments in 2024 and the following process evaluation(s) to gather valuable feedback on the delivery of the assessments and student, teacher and parent experiences. The gained information and lessons learnt will be used to further enhance the delivery of the assessments and student, teacher and parent experiences in the following years (i.e. a "learning-by-doing approach").

Although the Dutch writing assessment will not be adaptive, it is planned to be marked automatically from May 2025 (at the earliest). The necessary machine learning adds a level of complexity. In addition, the validity and reliability of machine learning marking should be closely monitored when implementing this innovative feature. Machine learning auto-marking sometimes relies on an analysis of mathematically based textual features (e.g. vocabulary/sentence length and complexity) rather than targeting deeper compositional type skills (e.g. creativity and persuasiveness) (Shermis and Burstein, 2013[115]; Ifenthaler, 2022[116]; Perelman, 2017[117]; 2020[118]). This threatens validity and could lead to a washback effect on teaching and learning surface skills if the stakes of the test grow. To reduce the complexity of the reform, the DoET could consider further delaying the auto-marking of the Dutch writing assessment, i.e. once the student assessments have been well-established.

Securing equity and fairness in the application of standardised student assessments

Reliability and validity are necessary conditions for any student assessment, and one cannot blindly assume that these conditions are met or transferable to all different subgroups of the population. It is important that assessments allow all students to show what they know and can do without being unfairly hampered by individual characteristics such as gender that are not relevant to what is being assessed (Reardon et al., 2018[119]; Binkley et al., 2011[120]). Student assessments need to be appropriate and sensitive to the needs of particular groups of students, such as students whose mother tongue is not the language of instruction and students with special educational needs (OECD, 2013[6]).

As discussed in Section 1, the standardised student assessments are intended to be as inclusive as possible. The draft decree specifically refers to two student groups for this, i.e. students with special education needs and non-Dutch speaking newcomers. Students in special education are not obliged to participate, with the exception of students in training path 4 of secondary education ("buso OV4"). However, these students can also be exempt if the school council so decides for a particular reason. Reasonable accommodations will be organised to allow these students to participate in the assessments. Students entitled to adaptations or educational materials during the school year may retain these when taking the student assessments (Flemish Government, 2023[53]).

The Steunpunt is taking these issues into account in the design and planning for the administration of the student assessments. For example, psychometricians from the Steunpunt are using the data from the large-scale piloting in 2023 to conduct differential function item analyses to ensure that test questions are fair towards different groups of students. The DoET should use the planned process evaluations and other research to examine this equity dimension of the standardised student assessments. The research findings may suggest further actions for enhancing the design and/or administration of the student assessments.

Use of student assessment results

Leaving school teams to decide on the use and weight of assessment results

The draft decree states that the class council can decide to use the student assessment results in the student's final marks. However, the standardised student assessment results cannot be used as the sole criterion for students' final marks for that subject (Flemish Government, 2023[53]). Schools must make agreements about this with teachers, students and parents via the school council and include these in their school regulations (Flemish Government, 2022[48]). The OECD has some concerns about leaving school teams to decide on the use and weight of the standardised assessment results for summative purposes. This might lead to differences between schools in terms of how they use the standardised assessment results in students' final marks. For example, where some schools might decide to use the standardised student assessment results for only a diagnostic/formative purpose, others might decide that the assessment counts as much as 30% of students' final mark. In the latter case, the student assessments

have stakes for students and even potentially for teachers in schools if they are held accountable, even if unofficially, for student results. The rationale behind this regulation is based on the Flemish tradition that devotes much value to the freedom of education and school autonomy, as well as the assumption that the school and the class council possess the necessary expertise and skills in student assessments.

The implicit prioritisation suggested by the test design decisions (for mathematics) suggests that using these results to support student learning at an individual level is not the primary (or a high priority) purpose of the student assessments. Therefore, following the examples of countries such as France and Wales (see Box 6), consideration should be given to using the assessment results for a mere diagnostic/formative purpose at the student level – at least in the initial years of implementing the standardised assessment. This would provide schools with some additional time to familiarise themselves with the data and feedback reports and respond through quality assurance processes for identified areas for improvement, including by investing in teachers' professional learning before schools can use the data as part of students' final mark.

This is particularly important considering the identified need for enhancing teachers' data literacy skills and inquisitive mindset to effectively enhance student learning and school leaders' capacity for using data in quality assurance and school development processes (OECD, 2019[121]; 2021[4]). Some of these needs were reflected in TALIS 2018, where, for example, 40% of lower secondary school principals reported needing further training for using data to improve the quality of their school (compared to 24% on average across the OECD). An earlier OECD study found that continuous professional learning in the Flemish Community tends not to involve teachers using data (OECD, 2021[4]). Also, while some providers offer continuous professional learning for teachers on data-informed practice, interviews indicated that few schools ask for the results of standardised student assessments and many consider that engaging with data (or research) remains a challenge for many teachers. Introducing the new standardised student assessments presents an opportunity to strengthen teachers' and school leaders' skills to effectively engage with the data from the standardised student assessments – as well as other assessment data for improving student learning (OECD, 2021[4]).

Section 3 will elaborate on the need to strengthen teachers' and school leaders' capacity, but if the Flemish Community continues leaving school teams to decide on the use of assessment results for summative purposes, the DoET should carefully monitor for possible unintended consequences.

Box 6. The use of standardised student assessment feedback reports for student learning in France and Wales (United Kingdom)

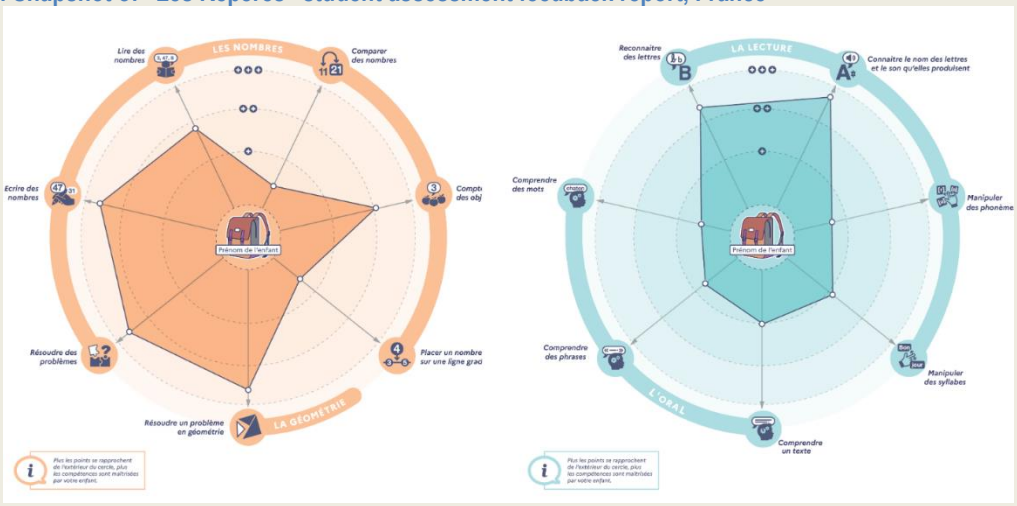
France

France has a range of statutory formative student assessments in primary and secondary education. These include "Les Repères", which consist of three standardised student assessments that are taken twice in "CP" (i.e. preparatory class, age 6-7) and once in CE1 (i.e. elementary class, age 7-8) by all students. Two are French language assessments and are devoted to mathematics; each takes about 20 minutes to complete. The assessments are important tools that enable teachers, students and their parents to monitor students' learning against national standards and adapt teaching and learning accordingly.

To support the use of the assessment results, teachers and parents receive two student feedback reports – one per subject – that show students' performance on each of the tested competencies (e.g. problem solving, understanding words, etc.), together with information on the assessments that clarifies their objective and implementation process. These data are visualised (among others) in a radar model (see Figure 9) to easily identify competencies of particular attention, as well as to provide an overall vision of the student's performance. The feedback reports include information to encourage parents to support their children's learning.

Teachers not only have access to the feedback reports but also to the students' assessments and a results guide. This guide provides a clear and detailed explanation of each competence, including how it is conceptualised, which items measure it, what students at each competency level are expected to know and do, and struggle with, etc. In addition, online resources provide teachers with guidance and ideas to adapt their teaching practices depending on their students' needs (French Ministry of National Education and Youth, n.d.[122]).

Figure 9. Snapshot of "Les Repères" student assessment feedback report, France



Wales (United Kingdom)

Wales has developed "personalised assessments" in reading and numeracy for learners in Years 2-9 (ages 7-14). These online, adaptive assessments are available to take throughout the school year. Their purpose is to help support progression in reading and numeracy skills. After learners have completed these assessments, schools have access to feedback on learners' skills, progress, age-standardised scores, and a range of reports to help plan teaching and learning.

The assessments – numeracy (procedural and reasoning) and reading (in Welsh and English) – are statutory for all learners in Years 2-9 (age 7-14) and have been phased in over a period of four years to replace the previous paper-based testing. They are adaptive in format and available for schools to use at any point in the school year that they consider most useful to provide feedback on skills and progress. Teachers and learners access them via "Hwb", the Welsh government's digital learning platform. The reports include feedback on individual learners' skills and progress, plus group reports showing skills profiles and progress charts which teachers can tailor for classes, year groups or bespoke groups.

The *individual learner feedback report* is first available to the teacher, who has the option to release the feedback to the learner when they consider it appropriate. This report provides a snapshot of a learner's skills at the time the assessment was taken. The first section, "The hardest questions I got right were on these skills", lists the areas in which the learner gave correct answers for the most challenging questions. The second section, "Some of the questions I got wrong were on", relates to the easiest questions answered incorrectly. The live links in the report link to a sample pdf for each of these questions. The final two sections relate to overall performance and provide an overview of the sort of skills that learners with a similar pattern of answers are able to achieve or are working towards.

Figure 10. Snapshots of the "personalised assessments" individual learner feedback report, Wales

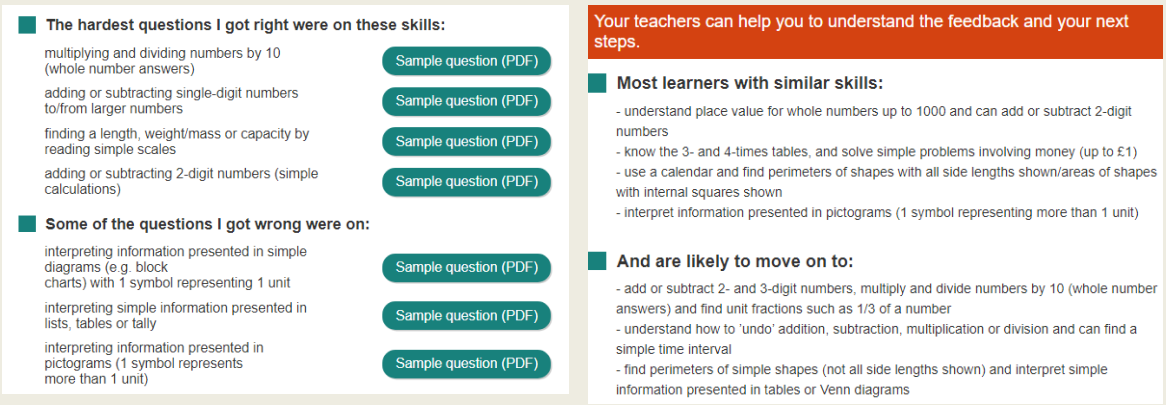
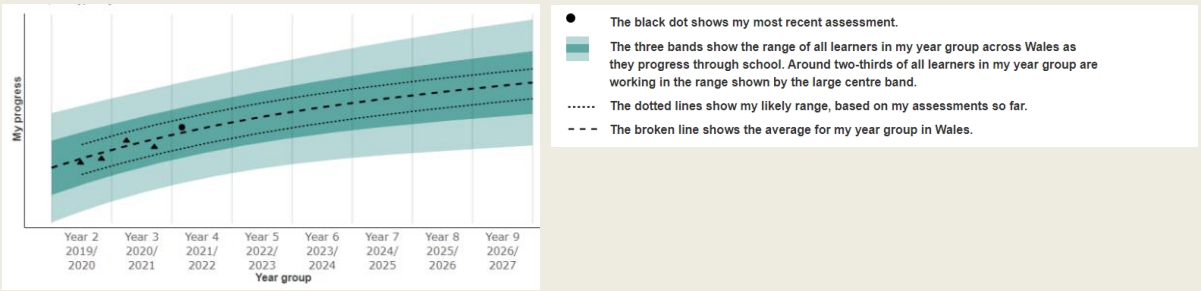


Figure 11. Snapshot of the "personalised assessments" learner progress report, Wales



A second report, the *learner progress report*, shows individual learner progress and contains a chart showing all the assessments taken to date in that particular subject. Some years may have more than one assessment shown, as schools can opt to run the assessments up to twice a year. The elements of the chart are carefully explained with simple language to facilitate understanding for all teachers, including those without a statistical or mathematical background. Schools are required to share both these reports with parents/carers. The reports include links for those seeking further information.

The Welsh government has developed a range of resources for schools, including a personalised assessment administration handbook, school case studies that highlight the use of personalised assessments to target and improve learner outcomes, webinars for teachers, and targeted information for parents and guardians to support awareness raising and actual use of the student assessment results for improving teaching and student learning.

Sources: French Ministry of National Education and Youth (2022[123]), *Évaluation Rèperes. Guide des scores. Notice technique*, https://eduscol.education.fr/document/41950/download; Welsh Government (2019[124]), https://hwb.gov.wales/curriculum-for-wales/reading-and-numeracy-assessments/personalised-assessments; Welsh Government (2022[125]), Personalised assessments. Guidance for practitioners on individual learner feedback and progress reports, https://hwb.gov.wales/api/storage/9ec62e0c-80d9-48c6-95c7-22f681f87c9a/personalised-assessments-guidance-for-practitioners-learner-feedback-sharing-22.pdf. Accessed on 25 October 2022.

Consider sequencing the use of student assessment results at different levels of the system

Given the multiple purposes of the standardised student assessments, their ambitious and innovative design and delivery, and – importantly – the reported challenges of teachers' and school leaders' data literacy skills and inquisitive mindset and consequent use of data for internal quality assurance processes (see above), stakeholders (i.e. schools, teachers and students) may benefit from sequencing the data uses at different levels of the system.

Different scenarios could be explored. However, recognising that the full cohort standardised student assessments are a vital source of information about the quality of education across the Flemish school system – i.e. for the first time, they will provide a standardised comparative measure of student performance across all schools in the Flemish Community, the immediate use of these results for system-level monitoring would seem warranted.

Also, as discussed above, consideration could be given to using the assessment results for a mere diagnostic/formative purpose at the student level – at least in the initial years of implementing the student assessment. This would provide schools with some additional time to familiarise themselves with the data and feedback reports, respond to identified areas for improvement, and invest in professional learning before the possible use of these data as part of students' final marks.

Building on these suggestions and the inputs provided by several education stakeholders during the interviews with the OECD, one option could be to have the immediate use of the assessment results for system-level monitoring and internal quality assurance by schools from 2024 onwards, to be followed by the use of the assessment results in external quality assurance by the Inspectorate of Education in the following year(s). This would leave schools some time to familiarise themselves with the assessment, the resulting data and feedback reports, and to work with PDB and other possible partners to respond to the identified areas for improvement (through quality assurance processes). In addition, this option would allow using the time to invest in developing teachers' and school leaders' skills (i.e. in data literacy, evidence-informed school development, etc.) before the data are used by the Inspectorate of Education.

That said, recognising that these student assessments are a vital source of information about the quality of education across the Flemish school system, the Inspectorate of Education may still want to directly trial the use of these assessment results for its new approach to prior screening (i.e. risk-based analysis) that is part of the differentiated approach to inspections under development. The release of data to research institutions should arguably occur directly in 2024 to allow validation and other research activities to take place (see Section 4).

Ensuring quality and timely feedback on assessments results to support improvements

Providing quality and timely feedback is one of the most effective teaching and learning strategies and has an immediate impact on learning progress (Dumont, Istance and Benavides, 2011[126]; Hattie, 2008[127]). Feedback reports are arguably also one of the most important elements of student assessments for supporting improvement in teaching and student learning (Somers, 2021[67]). The draft decree includes the intent for feedback at four levels: the system-, school-, class- and student levels (see Section 1).

In line with earlier OECD recommendations (Shewbridge and Köster, 2021[5]), the Steunpunt has been engaging with schools to design and develop the feedback reports. In addition, the digital feedback reports (under development at the time of writing) are aimed at supporting the rapid delivery of feedback, which is expected to positively impact on the perceived value of the student assessments and their relevance for educators. The feedback reports are planned to display students' results mainly in a numerical, graphical form with graphs and tables. One option to explore is complementing these data and information with qualitative descriptive results and additional resources for students, parents and teachers that could aid the understanding and implications of results. This would seem particularly relevant for the Flemish

context, considering the reported challenges of teachers' and school leaders' data literacy skills and inquisitive mindset, and the consequent use of data for internal quality assurance processes (OECD, 2021[4]; Shewbridge and Köster, 2021[5]). The Flemish Community may look towards the feedback reports and additional resources used in France and Wales as a source of inspiration (see Box 6 (OECD, 2013[6])).

Recommendations

- **There is a need to clarify the main purpose or purposes of the Flemish standardised student assessment**. The test design decisions taken for the mathematics test suggest an (explicit or implicit) prioritisation, or possibly a less systematic approach. **This is an issue for all partners involved, i.e. the Steunpunt, the Steering Group, the ministry and the Ministerial Cabinet, to clarify. If a prioritisation of purposes has occurred, this should be communicated widely.**

  On the other hand, **if no prioritisation has taken place, the DoET should conduct a prioritisation or weighting exercise of test purposes with all stakeholders to inform the ongoing and future design efforts and uses of the tests.**

  In addition, given the importance of the implications of the technical decisions and to avoid unintended consequences, **there seems to be a need to clarify the roles and responsibilities of the Steering Group and strengthen the oversight of the Steunpunt by directly involving the ministry and the Ministerial Cabinet in decisions on key issues**. For such issues that are of particular importance, like the proposed prioritisation of purposes, the Steunpunt could offer different scenarios, explaining their respective strengths, weaknesses and implications for the Steering Group and the ministry and Ministerial Cabinet to decide on.

- The DoET and/or the Steering Group may want to reconsider not sharing the assessment questions with teachers to allow them to explore in detail which type of items students succeed or struggle with. **If the assessment questions are not released, it should ensure schools and teachers still receive adequate and accessible feedback and support for interpreting the student assessment results**.

- Given the complexity of the student assessment reform and necessary management of expectations, **there seems to be much to gain from clarifying and communicating the possibilities and limitations of using the data for measuring learning gains and school value-added, especially in the initial years**.

  - **The Steunpunt should – as planned – engage with a broad range of stakeholders to develop the Flemish definition and measurement of valued-added school performance.** It should define the variables to be taken into account to obtain "fair" school comparisons. It may look towards the examples of other countries and the earlier undertaken feasibility study on the Flemish standardised student assessments to identify the strengths and weaknesses of different definitions and measurements, and work towards valued-added measures fitting the Flemish context.
  - **The proposed definition of value-added student performance should be agreed upon by the key partners involved, i.e. not only the Steering Group, but also the ministry and Ministerial Cabinet, to ensure all political and social dimensions are considered.**

- **The DoET should ensure continuous alignment between the student assessment content and the attainment targets over time.**

- There is value in evolving standardised student assessments over time based on concrete experiences; **a "learning-by-doing approach" may benefit the successful implementation of the student assessments**.

  - **The large-scale piloting will also provide valuable insight into the challenges schools may be facing in administering the assessments, as well as showcase what went well – both providing valuable lessons for their administration in the coming years.**

  - **It will be important to communicate beforehand that the administration of the assessments initially may not go perfectly, noting that this is to be expected.** Lessons learnt from this "learning experience" will be

taken forward in the following rounds of the assessments, pointing to the process evaluations planned for the piloting and first year(s) of implementation.

- Given that the large-scale piloting in May 2023 is planned without the adaptative testing feature, **the DoET should – as planned – use the first administration of the adaptive assessments in 2024 and process evaluation(s) to gather valuable feedback on the delivery of the assessments and student, teacher and parent experiences.** In addition, to reduce the complexity of the reform, the DoET could consider further delaying the automatic marking of the Dutch writing test to later, i.e. once the student assessments become well-established.

- **The DoET should use the planned process evaluations and other research to examine this equity dimension of the design and administration of the standardised student assessments.** The research findings may suggest further actions for enhancing these.

- Leaving school teams to decide on the use and weight of the standardised assessment results for summative purposes as part of students' final marks may lead to differences between schools. **Consideration should be given to using the assessment results for a mere diagnostic/formative purpose at the student level – at least in the initial years of implementation.** This would provide schools with some additional time to familiarise themselves with the data and feedback reports and respond to them through quality assurance processes for identified areas for improvement, including by investing in the professional learning of teachers before schools can use the data as part of students' final mark.

  If the Flemish Community continues leaving school teams to decide on using assessment results for summative purposes, the DoET should carefully monitor for possible unintended consequences.

- **Recognising the reported challenges of teachers' and school leaders' data literacy skills and inquisitive mindset and consequent use of data for internal quality assurance processes, stakeholders (i.e. schools, teachers, students and others) may benefit from sequencing the data uses at different levels of the system.** One scenario could be to have the immediate use of the assessment results for system-level monitoring and internal quality assurance by schools from 2024 onwards be followed by the use of the assessment results in external quality assurance by the Inspectorate of Education in the following year(s).

  Recognising that these student assessments are a vital source of information about the quality of education across the Flemish school system, the Inspectorate of Education may still want to directly trial the use of these assessment results for its new approach to prior screening (i.e. risk-based analysis) that is part of the differentiated approach to inspections under development.

  As noted above, consideration should be given to using the assessment results for a mere diagnostic/formative purpose at the student level – at least in the initial years of implementing the student assessment. The release of data to research institutions should arguably occur directly in 2024 to allow validation and other research activities to take place.

3. Ensuring strong stakeholder engagement, communication and a conducive context: Conditions for successful implementation

Building on the analysis in Section 2, this section starts by looking at the involvement of and communication with stakeholders in the design and implementation of the standardised student assessments, which research evidence shows is a key means for ensuring the relevance and quality of a new policy and gaining the much-needed support and ownership for its successful implementation. This is followed by an examination of other key factors of influence for the "conduciveness" of the environment in which the standardised student assessment reform is implemented, such as the clarity of roles, responsibilities and expectations among implementing partners and other key stakeholders, including a discussion on possible governance structures to deliver the student assessments in the longer term. The section concludes with a discussion on developing teachers' and school leaders' (data literacy) skills to effectively use data for improving teaching and learning.

Stakeholder engagement and communication

Continuing and further enhancing stakeholder engagement and "co-construction" of the reform

Research evidence has highlighted the importance of combining top-down and bottom-up strategies, following a participatory approach, to successfully navigate the complexity of education policy design and implementation (Blomkamp, 2021[128]; OECD, 2020[129]; 2020[8]; Nogueira and Schmidt, 2022[130]). Increasingly, OECD countries are recognising the importance of engaging stakeholders early in the policy design stage as a key means for ensuring the relevance and quality of the new policy and gaining the much-needed support and ownership for its successful implementation (Viennet and Pont, 2017[9]; Burns and Köster, 2016[10]; Schleicher, 2018[12]). These meaningful participatory practices have been gaining momentum over the last decade (OECD, 2020[7]; Torfing, Sørensen and Røiseland, 2016[131]). The Flemish education system benefits from a culture geared towards stakeholder engagement and co-construction of policies.

An earlier OECD report (Shewbridge and Köster, 2021[5]) noted the potential of the high-level forum (established in 2021) as a specific stakeholder consultation platform to facilitate communication and feedback at key stages of the development of standardised assessments (see Box 1). Stakeholders include representatives of the umbrella organisations, teacher unions, the Inspectorate of Education, the Steunpunt and the DoET. The high-level forum has been mandated to communicate and consult on the

key decisions in policy development for introducing the standardised assessments. Its objectives are threefold: 1) the global monitoring of the progress of the reform; 2) providing a space for stakeholders' feedback and inputs; and 3) providing information on the reform, including timing and next steps (Flemish Ministry of Education and Training, 2021[132]). At the meetings of the forum, stakeholders are given information on recent policy development (Shewbridge and Köster, 2021[5]). The forum met five times between May 2021 and February 2022, bringing together around 30 stakeholders in each meeting. Despite the initial progress, the stakeholder engagement process through the high-level forum has not yet realised its potential, as evidenced by the fact that it has not met since February 2022. Discussions on the format and role of the high-level forum, with some stakeholders preferring more formal negotiations, have remained unresolved.

The previous OECD study also noted the importance of committing to stakeholder involvement and ensuring key voices other than national level education stakeholders are heard at every stage of the policy development process (Shewbridge and Köster, 2021[5]). These key voices include school leaders, teachers and parents. In particular, it noted the potential of raising awareness and mobilising support and feedback channels for school leaders. Apart from the engagement of some school leaders in the design of the student assessments and feedback reports (see below) – a strength of the reform – so far, the reform has failed to engage school leaders at scale. Arguably, it is not only a matter of scale but about engagement in a broader sense, providing school leaders with the information, power and capacity to be key actors in the reform. The limited communication on the reform has hampered the much-needed mobilisation of awareness, support and feedback from school leaders, teachers, students and other stakeholders, as will be elaborated below.

The previous OECD study also highlighted the importance of organising contributions from the educational field to support the Steunpunt's work (Shewbridge and Köster, 2021[5]). The Steunpunt (as mentioned earlier) is a consortium of all five Flemish universities and two universities of applied sciences commissioned by the Flemish government to support the introduction of the standardised student assessments during the period 2021-25. It brings together Flemish researchers with the necessary expertise to design the computer-based and (partially) adaptive standardised student assessments, and support schools in administering the assessments, thereby "pooling" the available expertise and capacity in the Flemish Community for this highly technical work. Importantly, the Steunpunt consults and collaborates closely with the educational field (see Box 7). In its first year of operation, the Steunpunt had already engaged with more than 400 stakeholders working in schools, the Inspectorate of Education and the PDB. It has established working groups for designing and developing the Dutch language and mathematics student assessments and focus groups with teachers and school leaders to explore the desirability of reasonable adjustments for students that need these (Steunpunt, 2022[133]).

Steunpunt researchers have also started monitoring teachers' and school leaders' views, expectations and readiness for the student assessments (Molenberghs et al., 2022[54]). The findings of such research could greatly benefit the implementation of the student assessments. For example, failing to identify negative emotions towards the student assessments would hinder its successful implementation (Rafferty and Minbashian, 2018[134]; Kin et al., 2015[135]; Armenakis et al., 2007[136]). On the other hand, early identification of such views could allow modifying the reform strategy, for example, by expanding and further targeting communication messages to explain to teachers and school leaders the rationale and potential benefits of the assessments for improving their practice.

The Steunpunt has further engaged the education profession in the layout of the feedback reports and the identification of professional learning needs. The continued engagement with school leaders and teachers will be vital for further enhancing and tailoring these to their needs; thereby further enhancing their capacity to use the assessment results to make sustainable improvements in teaching and learning.

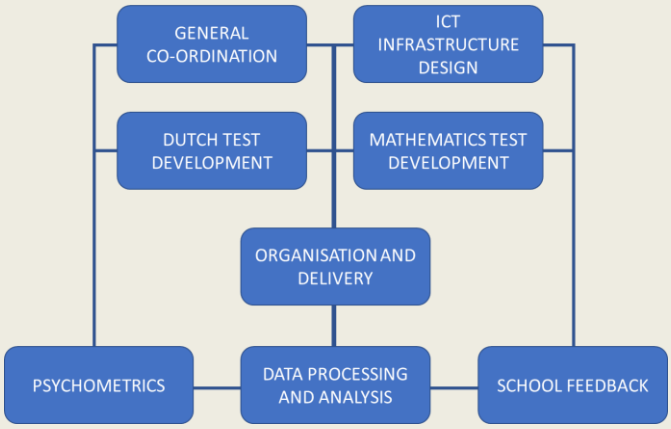
Box 7. The Steunpunt central tests: An example of stakeholder engagement and "co-construction" and pooling of expertise in test design and administration

The Steunpunt central tests, or "Steunpunt", consist of a consortium of all five Flemish universities (Ghent University, Vrije Universiteit Brussel, Katholieke Universiteit Leuven, Hasselt University and Antwerp University) and two universities of applied sciences (Artesis Plantijn Hogeschool and Arteveldehogeschool). The Flemish government established a partnership with the Steunpunt to support the design and implementation of the Flemish standardised assessments between 2021 and 2025. The Steunpunt has five core tasks:

1. The first core task is the development of the test items based on the attainment targets. The Steunpunt focuses on developing meaningful assessments that challenge students to demonstrate their skills in mathematics and Dutch.
2. For the second core task, which is to support schools in administering the assessments, the Steunpunt develops accessible scripts and online courses, among other things. Schools can also count on the availability of a helpdesk, where they can go with questions about the assessments.
3. After administering the assessments, the Steunpunt processes and analyses the test results. Based on these results, the Steunpunt then provides feedback to the educational field.
4. For the fourth core task, the Steunpunt is committed to developing accessible online courses and feedback reports (dashboards) for the educational field.
5. The final core task concerns the development of expertise and research into educational effectiveness, for example, into which school factors make the difference for strong education.

The core tasks are performed by a multidisciplinary team composed of subject matter experts in Dutch and mathematics, teachers (trainers), psychometricians, pedagogues and data analysts. The multidisciplinary team works in eight sub-teams or work domains (Figure 12).

Figure 12. Structure of the Steunpunt's work domains



To complete its five core tasks, the Steunpunt follows four related principles, which can be shortly described as aiming for meaningful, valid, reliable and accessible assessments that are computer-based and adaptive in nature. The Steunpunt works closely with various stakeholders in the educational field. Active stakeholder engagement and co-construction take place in the form of think tanks, test trials and working groups. During its first year of work, the Steunpunt engaged more than 400 stakeholders and continues to work closely with the education profession to carry out the large-scale piloting in 2023, and onwards.

In sum, although the OECD identified a need to enhance the oversight of the Steunpunt by directly involving the ministry and Ministerial Cabinet in taking decisions on key issues and proposals (see Section 2), the Steunpunt is a good example of stakeholder engagement and co-construction. This approach should be continued, and possibly expanded on as the reform advances from a focus on assessment design to one

more focused on the administration, analysis and reporting, and actual use of the assessment results in schools.

Limited communication on the student assessment reform

Effective communication in education is paramount for stakeholder engagement, and for successful and sustainable policy making and implementation. Clear communication can enhance stakeholder engagement and help mobilise support for the reform (Threlfall and Althaus, 2021[137]; Althaus et al., 2021[138]; Ansell, Sørensen and Torfing, 2017[139]; Schleicher, 2018[12]). A new policy or reform is unlikely to succeed unless those expected to implement it see its value, want to see it happen, are confident in their capacity to implement it and, ideally, have ownership of the change (OECD, 2020[8]; McKnight and Glennie, 2019[11]; May, 2015[51]). The earlier OECD study on the standardised student assessment reform also identified the importance of investing in clear and active communication as a priority for the DoET (see Box 1) (Shewbridge and Köster, 2021[5]). The OECD had to conclude, however, that this remains an area for improvement, as will be elaborated on below.

At the start of the reform, the DoET initiated a number of important initiatives to share information regarding the Flemish student assessments with various audiences. As previously explained, this included the establishment of the high-level forum that was created, among other reasons, as a space for information sharing and receiving feedback on policy developments. The DoET also developed a dedicated webpage on its official website that captures key information on the assessments (Flemish Ministry of Education and Training, 2022[140]). It has also organised and participated in several events to share information with various education stakeholders (e.g. teachers, school leaders, etc.), such as a public online Q&A session ("*vragenuurtje*") in February 2022 (Flemish Ministry of Education and Training, 2022[140]; 2022[141]). The OECD was invited to listen in to this event and was impressed with the questions and open discussion between the DoET officers and the participants, many of whom were teachers and school leaders.

However, these initiatives were relatively limited in number and outreach. For example, in the high-level forum meetings, as with any such forum, only a limited number of stakeholders could participate. Furthermore, the initial emphasis and efforts placed on communication were reduced during much of 2022 due to the political sensitivities surrounding the reform at the time.

Therefore, while national level stakeholders, like the umbrella organisations, teacher unions and the Inspectorate of Education, are all informed about the reform, there is much work to be done to raise awareness and understanding of the reform among the education profession (Molenberghs et al., 2022[54]) as well as students, parents and the general public. International experiences in introducing student assessments show that it is vital for school leaders, teachers, students and parents to be informed of the rationale for introducing student assessments – why they are implemented, their purpose, what students can expect from the assessment experience, etc. (OECD, 2013[6]; ACER and UNESCO Institute for Statistics, 2017[142]). The reform narrative and key messages should also aim to clarify why the assessments are needed and, in particular, how they can be used to improve the quality of education.

Regarding the key messages, several stakeholders raised in the interviews the need for the DoET to clarify and expand on its key message that "with the Flemish student assessments the Flemish government wants to strengthen the quality of education" (Flemish Government, 2022[48]; Flemish Ministry of Education and Training, 2022[140]). Although seemingly not questioning the rationale of the reform, which points to the decreasing student results, various stakeholders noted that in the communication to date, there was an insufficient explanation of the mechanisms by which the assessments would improve educational outcomes (the "how").

The generally limited communication on the standardised student assessment reform and the fact that the large-scale piloting of the assessments was scheduled for May 2023 with the first administration the following year, argue for urgency in the development and operationalisation of a comprehensive

communication strategy. This strategy should explain the "why" and the "how" the standardised student assessments – together with other policy measures – can support the realisation of the overarching goal of improving the quality of education. For example, the reform narrative and key messages could focus on answering questions such as "How will the schools use the data to support improvements in student learning?"; "How will the Inspectorate of Education use these data?"; "How do the standardised student assessments complement and add to already existing student assessments?".

To help further clarify the rationale (i.e. the "why") of the reform, part of the communication could highlight the fact that (as mentioned above) the Flemish Community does not have central exams at the end of compulsory education nor full cohort standardised student assessments in support of system-level monitoring and improvement efforts at different levels of the system, which makes it an outlier among OECD countries (OECD, 2013[6]; 2015[25]; Shewbridge, Fuster and Rouw, 2019[23]).

Several stakeholders also noted the need to link the reform to the broader policy environment, pointing specifically to the need to invest in teachers' continuous professional learning, including their data literacy skills. Ensuring teachers and school leaders have the skills to use the assessment results effectively will need to be considered in their continuing professional learning (OECD, 2021[4]). This suggests that the narrative needs to set out how the introduction of standardised assessments is one of a range of important policy measures, like those on continuous professional learning and staff shortages, to help improve the quality of education in all schools in the Flemish Community of Belgium.

In addition, recognising the equity challenges of the Flemish school system (see Section 1), the overarching goal of improving and monitoring the quality of education could also be expanded to emphasise the need for improving and monitoring equity in education. The standardised student assessments will introduce reliable and objective student performance results across all schools in the Flemish Community, allowing the highest levels of needs and disadvantage to be identified and subsequent support for those schools.

Furthermore, in light of the Flemish education culture, it would seem important to reiterate the message that the standardised student assessments are aimed at serving the primary goal of supporting improvement (Flemish Government, 2022[48]), i.e. offer an important opportunity for "learning" and supporting all schools, school leaders and teachers in innovating their practice and helping them realise sustainable improvements in student learning.

The communication strategy should ensure that messages reach all stakeholders (i.e. teachers, school leaders, ICT co-ordinators and support staff in schools, students, parents, etc.) so that they are clear on the reform rationale, what it entails for them and the system, leading to appropriate expectations. For this, the DoET should work with the education profession, PDB, the Inspectorate of Education and other partners to develop a coherent narrative that is communicated to all. The DoET and education stakeholders should collaborate to ensure consistent messaging; optimise reach; and use multiple media channels based on reach, frequency and credibility (see Box 8) (Cabañero-Verzosa, 2009[143]).

Different stakeholders may also need targeted forms and levels of information (Behrstock-Sherratt, Biggers and Fetters, 2012[144]). For example, the adaptive nature of (part of) the student assessments, measurement of learning gains and the planned use of contextualised school results are complex issues that should be well explained in an easy-to-understand manner. Given the diversity of the stakeholders involved, the messages should be tailored to each audience and use different communication means (Viennet and Pont, 2017[9]). For example, regularly using communication channels such as *Klasse*, a popular, well-established education magazine in the Flemish Community, could maximise reach among teachers. Also, a simple guide and/or animation could be circulated to explain the assessment procedure to students.

Box 8. Colorado's communication and engagement strategy for evaluating teachers and principals

In 2013, the state of Colorado (United States) implemented a statewide reform introducing a new system for evaluating teachers and principals. The state invested significant time and resources to develop a communication and engagement strategy to support the implementation of the reform, hiring a communication expert to assist with this task. The National Comprehensive Center for Teacher Quality selected Colorado's strategy as an interesting case study, given the variety of successful strategies implemented and their emphasis on transparency. Below are some examples of the initiatives carried out by the Colorado Department of Education (CDE).

- **"Staying Informed":** The CDE created an openly accessible document which outlined how to stay informed about the reform, explaining all the means and channels by which the CDE would be sharing information and reaching out to stakeholders and the community more broadly.
- **Interdepartmental brown-bag lunches:** The team responsible for implementing the reform organised brown-bag lunches to share information and discuss a common approach for alignment and communication within the CDE. This resulted in well-informed staff, a more collaborative environment and greater coherence across initiatives. Ultimately, it led to consistent and coherent messages delivered to stakeholders, avoiding misinformation.
- **Events:** Four symposia were conveyed in different locations around the state to communicate on the legislative changes and recruit pilot districts. They were so successful that the same procedure was repeated the following year. In addition, more than 500 participants attended a statewide one-day summit where evaluation policies, rules and timelines of the reform were presented, and spaces to enhance collaboration between district teams and experts were provided.
- **Distribution of user-friendly resources:** The CDE distributed resources with practical information targeted to different stakeholders. These included:
  - An Educator Evaluation System Implementation Toolkit, which provided guidance on how to prepare for implementation and a worksheet to self-assess a district's readiness for compliance with the new law.
  - Two district leaders guides: One focused on communicating effectively with stakeholders and the other on preparing districts to implement the new educator evaluation system. The first provided concrete advice and ideas, and the second an overview of the first steps leaders should take (e.g. establishing goals, assembling a design team), both including lessons learnt from other districts.
  - A resource bank webpage hosting all relevant information and resources to implement the reform successfully, such as links to the state's model evaluation system, sample evaluations, and wider resources from national experts and technical assistance providers.
- **Written communications initiatives:**
  - Dedicated website: A webpage specifically dedicated to the reform and regularly updated was made public. It included regular reporting of meeting highlights, a frequently asked question section, links to related initiatives and all relevant public documents, including explanations comparing different versions of rule-making documents.
  - Newsletter: An e-newsletter was launched and regularly distributed to inform on progress and provide updates. Voluntary sign-up was open to all.

In addition to these initiatives, the CDE worked on establishing good relationships with stakeholders from the design of the reform and maintaining constant communication with different stakeholder groups throughout its implementation.

Source: Behrstock-Sherratt, Biggers and Fetters (2012[144]), Lessons Learned on Communication and Engagement for Educator Evaluation: Colorado Case Study, https://files.eric.ed.gov/fulltext/ED543667.pdf.

Celebrating success and recognising the introduction of any new policy is unlikely to go perfectly

As much as identifying gaps, needs and areas for improvement is key, it is also important to focus on the positive factors when communicating about new policies or programmes. The deliberate identification and celebration of successes and their public acknowledgement can reinforce the implementation effort. Failure to do so can have adverse effects. For example, ACARA's *Departmental Report* reflected on this issue by highlighting stakeholders' criticism on their communications on student assessments, which were often interpreted as defensive and reactive. There was a clear need to pursue a more positive and informative communication strategy, and share the value and celebrate successes (Australian Department of Education and Training, 2019[145]).

Drawing from this example, the communication strategy for the Flemish standardised student assessment reform should celebrate successes, as these may help raise awareness, support and ultimately favour the successful implementation of the reform. Such successes could, for example, include the high participation rates of schools and students, the positive experiences of students in taking the assessments, or of teachers using the results to help students in their learning.

Again, different media channels and tailored messages to different stakeholders could be mobilised to reach a broad audience. For example, case studies that describe the effective use of feedback reports in school development efforts can disseminate success stories and provide concrete guidance to schools that have identified similar improvement priorities. Such "scaffolding" based on schools' "improvement journeys" may be an effective means for mobilising and expanding the school development support capacity of the system.

At the same time, implementing a standardised student assessment, or any other new policy, is unlikely to happen without encountering some challenges. As mentioned earlier, it will be important to communicate beforehand that in the assessment administration, things may not initially go perfectly, noting that this is to be expected. From this "learning experience", lessons learnt will be taken forward in the following rounds of testing, pointing to the process evaluations that are already planned for the piloting and first year(s) of implementation (Flemish Government, 2022[48]).

To ensure stakeholders are an active part of this learning journey and incentivise a bottom-up approach, communication should be bidirectional (i.e. two-way communication) (OECD, 2020[129]; 2020[7]; Viennet and Pont, 2017[9]). Although the high-level forum was developed to create such a space for feedback and discussion, the chosen format limits the number of participants and, as mentioned, some stakeholders preferred more formal negotiations. Therefore, the DoET should consider developing additional two-way communication channels that create opportunities for a wide range of stakeholders, including school leaders, teachers and students, to provide feedback and comments beyond the introduction of the standardised student assessments. This could be in the form of focus groups or the establishment of an open digital portal.

In addition, the DoET should facilitate independent monitoring of the awareness, readiness and support for the assessments among teachers, school leaders, students and their parents. Governments are often seen as unresponsive to public feedback, which has an impact on trust and perceived reliability (OECD, 2022[146]); gathering feedback can only lead to improvement if it is acted upon. An appropriate, timely analysis and response to this feedback could help retain stakeholders' support and engagement, as well as identify implementation gaps and provide guidance on how to respond to these.

Roles and responsibilities

The need for ensuring clarity of roles, responsibilities and expectations

Clarity and transparency of roles and responsibilities foster trust among stakeholders, is a collective effort, and involves stakeholders in defining their roles and monitoring their performance. Such measures include having a clear task allocation, accountability and a monitoring system to gauge progress. Whereas task allocation sets boundaries and determines who is (or are jointly) in charge of what, accountability relates to responsibility and reporting relationships (OECD, 2020[8]).

A key means for clarifying roles, responsibilities and expectations is the development of official guidelines to support the implementation of the student assessments in schools (OECD, 2013[6]). The Steunpunt should, as planned, develop official guidelines that clearly describe the roles, responsibilities and expectations of everyone involved in the test administration (i.e. teachers, school leaders, ICT co-ordinators, students). The experiences from the small-scale piloting that was started in 2022 and the

large-scale piloting in May 2023 could inform the development of these guidelines based on the experiences of school leaders, teachers, ICT co-ordinators, students and others. The proposed communication strategy should help raise awareness of these guidelines and supporting resources (e.g. animators, videos), with key education stakeholders (e.g. the Steunpunt, umbrella organisations, teacher unions, the Inspectorate of Education, the DoET) playing vital roles in supporting their implementation.

Considering how to best govern the standardised student assessments in the medium- to longer term

Evidence highlights the importance of ensuring adequate system capacity for the design, implementation, reporting and feedback of results when introducing new full cohort student assessments (OECD, 2013[6]; Verger, Fontdevila and Parcerisa, 2019[147]; Reedy, 2019[148]). For the design and implementation of the standardised student assessment, the DoET commissioned the support of the Steunpunt for the period 2021-25. As mentioned earlier, this collaboration, among others, benefitted from the researchers' test design expertise and ensured engagement and trust among a wide variety of stakeholders in the Flemish Community.

When an assessment system becomes fully operational, public acceptability is strongly influenced by problem-free delivery, which relies on solid assessment operations and appropriate evaluation capacity. Despite the progress made to date with the Flemish standardised student assessments, international experience confirms that full cohort student assessments are a complex, multi-faceted and technical endeavour (OECD, 2013[6]), which may challenge the capacity of Ministries of Education that often are already handling many other responsibilities.

In response to this challenge, OECD countries have pursued different ways to ensure optimal delivery and governance of their student assessments. Many have established a dedicated specialised assessment agency to deliver and govern the student assessments (OECD, 2013[6]); others have maintained these responsibilities within their ministry. Although comparisons between countries and jurisdictions are challenging, among others, due to the differences in roles, responsibilities, legal constructs and legal systems, this chapter discusses some of the different approaches taken by countries to inform the Flemish DoET and education stakeholders in deciding on possible next steps.

On one side of the spectrum are those countries that have created the necessary "distance" between the ministry and the student assessments by engaging an autonomous assessment agency in the form of a private company or by establishing and/or expanding the mandate of existing public agencies. One example of the former is provided by the Netherlands, where the agency CITO (Central Institute for Test Development) used to have the task of developing and delivering the national assessments and exams. Recently, the Dutch government decided to allow other test providers on the market as long as they meet the national standards and requirements on national assessments and exams. The quality of the test providers is checked and guaranteed by the independent College for Assessments and Exams.

Several OECD countries have chosen to engage and/or establish an autonomous specialised public agency to administer the student assessments based on some form of delivery agreements. In Australia, the annual NAPLAN is conducted by ACARA. ACARA is an autonomous statutory authority whose work is directed by the Australian education ministers. Its responsibilities are not limited to assessment but also include curriculum development and reporting on schooling (ACARA, n.d.[149]).

In other OECD countries like Denmark, England (United Kingdom), Ireland and Latvia, specialised assessment agencies are not autonomous from the government. In England, the Standards and Testing Agency (STA) is an executive agency which operates within the Department for Education. Because the STA is not demonstrably independent from the government to manage the perceived risk of political

interference in test outcomes, the assessments are regulated by Ofqual, an autonomous, non-ministerial arm's-length body (see Box 9).

These international examples show that the division of roles and responsibilities among different assessment agencies and other bodies can be complex, with the possibility of partnering with external private suppliers (like in England and the Netherlands). When this is the case, it is still important to have the appropriate expertise within the agency to support robust procurement and contract management. Such a possible engagement with private suppliers suits the Flemish context less.

Although not urgent, the DoET should consider how to best govern and deliver the standardised student assessments in the medium- to longer term. The government must carefully consider whether to (in due time) opt for a dedicated autonomous public agency or establish this within the government structure against the possible advantages and disadvantages of these options.

A potential advantage of having a specialised autonomous agency is the development of specialised technical expertise within the agency, including development assessment capacity and data literacy skills that can support the Flemish school system. Student assessment is a highly technical matter and its design and implementation require expert capacity, which takes time to develop. Such expertise is particularly important when education systems are developing and introducing large-scale standardised student assessments – as is the case for the Flemish Community. In addition, a deep understanding of the role of student assessment in improving policy and practice is essential to develop policy, support the development of assessment tools, and ensure that assessment results are used appropriately for monitoring and supporting schools. International experience shows that specialised agencies can develop such capacity by engaging with each other and with the assessment community more widely, both nationally and internationally (OECD, 2013[6]).

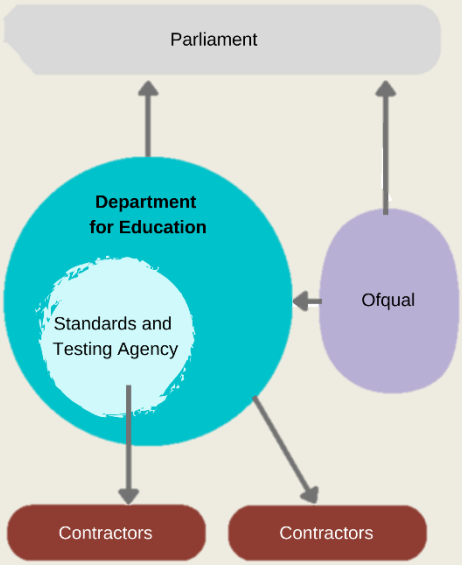
Furthermore, the design and implementation of standardised student assessments is a complex, multi-faceted endeavour that may weigh (heavily) on the capacity of the Ministry of Education, particularly when it is part of its direct roles and responsibilities and concentrated in its organisational structure. Establishing a specialised autonomous agency outside the ministry structure to take on this challenging task could reduce the burden placed on ministry officials. In addition, even though "full autonomy" is rare, as agencies often to some extent still depend financially on and are governed by the government, specialised agencies introduce some distance *vis-à-vis* education authorities in the potentially politically sensitive arena of student assessments.

On the other hand, establishing an autonomous agency can be more costly due to the costs of its initial establishment and longer term functioning. In some countries, however, dedicated agencies take on a broader role in monitoring and evaluating the quality of education, as well as other functions. For example, as mentioned above, ACARA is also responsible for developing the Australian National Curriculum and reporting on schooling. Consolidating several roles and responsibilities could make an agency more cost-effective while also facilitating coherence across different policy areas.

Box 9. Governance of the delivery of standardised assessments at the primary level in England (United Kingdom)

In England (United Kingdom), national assessments are delivered by the Standards and Testing Agency (STA). The STA is an executive agency sponsored by the Department for Education. The STA operates within the Department for Education (i.e. drawing on Department for Education staff) and the Secretary of State for Education is accountable to parliament on all matters concerning the agency. The STA's chief executive has a specific, fully delegated role with respect to test standards and standard setting and maintenance. The chief executive signs off on the content of the assessments and the setting and maintenance of standards, acting independently of the Department for Education and ministers to ensure confidence in the validity of test outcomes. However, because the STA is not demonstrably independent from the government, the assessments are regulated by Ofqual to manage the perceived risk of political interference in test outcomes. Ofqual is an independent, non-ministerial arm's-length body. It reports annually on its regulation of the assessments, particularly regarding their validity and the maintenance of standards of attainment (including over time). The underlying function here is transparency and public confidence in test outcomes. The STA outsources the delivery of the assessments and conducts some elements of test development and evaluation of outcomes in-house. For example, the test design and the psychometric analyses necessary to set and maintain standards are done internally, but the delivery of test papers and marking are outsourced to contractors with the requisite experience and resources. In addition, local authorities have a range of statutory responsibilities, such as monitoring test administration to ensure the integrity of the assessments.

Figure 13. Diagram of the governance structure of the national assessments in England



These arrangements have evolved, however. In the past, a non-departmental public body (the Qualifications and Curriculum Authority) was responsible for national testing. The assessments were delivered by a division of the Qualifications and Curriculum Authority – the National Assessment Agency. The National Assessment Agency contracted external suppliers to produce, distribute and mark the assessments, and collect the data for reporting purposes. Changes were made following issues with the timely delivery of results in 2008, with the Qualifications and Curriculum Development Agency taking responsibility for the assessments for a period, but the model of test development and distribution was maintained until such duties were finally transferred to the STA in October 2011. As illustrated in this example, arrangements and responsibilities may evolve to respond to newly identified needs and optimise the delivery of the assessments.

To conclude, internationally, there are different ways in which standardised assessments are delivered and governed. This is because any system involves complex interactions of history, culture, and educational policies and practices (Opposs et al., 2020[150]; Verger, Fontdevila and Parcerisa, 2019[147]). Whether to (in due time) opt for a specialised autonomous agency or establish this within the government structure is an issue for further reflection and, as mentioned, should consider the advantages and disadvantages of both options.

However, it would seem vital to build on the expertise available in the Steunpunt. Currently, the collaboration with the Steunpunt is established until 2025. By 2025, the researchers in the Steunpunt would not only hold knowledge and expertise but would also have vast hands-on experience of the Flemish test implementation. There are tentative plans to have a new Steunpunt starting in 2025, with a slightly different set of tasks. Such a measure could help maintain the momentum gained in implementing the reform. There, for example, would remain the necessary work to further develop and calibrate test items, measure learning gains and value-added performance, and evaluate whether schools interpret and respond to test feedback appropriately. Care should be taken to ensure the necessary operational and logistical expertise is present to facilitate the implementation complexities elaborated upon. Whether this is a temporary, intermediate measure or envisaged to be(come) a permanent structure is not clear at present.

Whatever construct is decided on, it should have strong governance and oversight, so reform objectives are achieved, policy and practice align, and risks are monitored and managed across all relevant bodies.

Capacity

Enhancing teachers' and school leaders' skills to effectively use data to improve teaching and learning

Research evidence shows the potential of using data to improve student learning (Schildkamp and Lai, 2012[151]; Airola and Dunn, 2011[152]; Faria et al., 2012[153]; Poortman and Schildkamp, 2016[154]). For this to happen, it is paramount that education professionals acquire data literacy skills, not only to interpret data appropriately but also to take advantage of all the information it can offer (Mandinach and Gummer, 2016[155]). This was also recognised in the Education Policy Note 2019-2024, in which the Flemish government announced the introduction of standardised student assessments (Weyts, 2019[156]). This note highlighted the importance of developing adequate data literacy skills at the school level to ensure that teachers can use assessment results to inform the improvement of teaching and learning processes. It considers the new student assessments an opportunity to professionalise teachers and school leaders in data literacy, emphasising its use for school development and enhancing student learning.

When introducing data literacy professional learning opportunities, it is important not only to do this in the context of the standardised student assessments but more broadly and focus as well on how to develop, mark and use self-developed formative and summative assessments. This could lower the risk of teachers over relying on the new standardised assessments, and questioning their own skills and priorities when assessing student's learning needs and performance (Osborn, 2006[157]). If teachers imitate the format and content of the standardised assessments when creating classroom formative and summative tests, the alignment between the intended, enacted and assessed curriculum could be jeopardised. This alignment is key, as it prompts teachers to reflect on fundamental questions, such as whether the assessments map onto the objectives of the intended and enacted curriculum and can improve teaching quality and learning outcomes (Pasquini and DeLuca, 2022[158]). For this reason, schools should encourage and support teachers to develop their own student assessments.

While internationally, the development of teachers' data literacy skills is in the spotlight, the data literacy skills of school leaders are also key for school and student success. School leaders have an important role

to play in setting a school's culture around data use and in promoting its use as part of internal quality assurance and development processes (Henderson and Corry, 2021[159]; OECD, 2019[121]). A school's culture of data use can have serious impacts on how teachers think about and use data (Farrell and Marsh, 2016[160]). As such, both teachers and school leaders will need to develop their data skills to effectively use the new standardised student assessment results to improve teaching and learning in their schools.

Clarifying the concept of data literacy and underlying skills

Although scholarly interpretations of the concept differ, in the field of education, "data literacy" is often defined as the ability to understand and use data effectively to inform decisions. It comprises a specific skill set and knowledge base that enables educators to transform data into information and ultimately into actionable knowledge (Mandinach and Gummer, 2016[155]). Data literacy incorporates aspects of statistical literacy, assessment literacy, pedagogical knowledge and data-driven decision making under one umbrella (Gummer and Mandinach, 2015[161]). This domain of professional knowledge comprises the ability to analyse and interpret raw and statistically, tabularly and graphically presented data and to translate those interpretations into instructional decisions, including the students/content on which to focus, instructional method(s) to use, and how to differentiate or modify teaching. In addition, data literacy involves specific analytical tasks, such as examining student work for patterns (e.g. common errors or misconceptions) and item-level data, making data-based comparisons, and disaggregating data.

Technology is not separate from data literacy. In many OECD countries, student information systems and online testing systems are just a few of many platforms that teachers need to learn to navigate, extract data and interpret visualisations. The role of technology in data-informed decision making will increase as machine learning and artificial intelligence systems make recommendations based on big data from student keystrokes, progress and outcomes tracked in learning systems (Henderson and Corry, 2021[159]).

Digitalisation in education has been gaining attention internationally with a focus, among others, on teacher professional development (OECD, forthcoming[162]). For example, Spain's 2022 strategy for digital education aims to improve teachers' digital competences. The large-scale initiative aims to train 80% of the country's teachers (about 700 000) in digital competences by 2024 (European Commission, 2022[163]). As mentioned earlier, the Flemish Community of Belgium is also investing in the digitalisation of the education sector. The "Digisprong" project includes investment in infrastructure (e.g. laptops, computers, software packages), Internet connectivity and teacher training. The project is scheduled for completion in 2024 (Government of Flanders, n.d.[106]; SGI Europe, 2021[107]). Through this and other initiatives, the Flemish Community aims to ensure teachers and school leaders gain the necessary ICT skills to support test administration, including supporting their students in practising with the test platform and sample questions and in accessing feedback reports on an online platform (that was under development at the time of writing).

Research suggests that one of the biggest challenges in moving forward with data literacy in pre-service teacher education and continuous professional development programmes is a lack of a common understanding of the construct and surrounding terminology, and the actual knowledge and skills involved. Even with a clear definition, there is a need to clarify the actual skills involved, which will be essential for developing skills-focused pre-service teacher education and continuous professional learning (Henderson and Corry, 2021[159]; Mandinach and Gummer, 2016[155]).

In light also of the fragmented continuous professional learning system in the Flemish Community (OECD, 2021[4]), the Flemish standardised student assessment reform may benefit from having the Steunpunt and other education stakeholders define the concept of data literacy and the specific skills it entails. Such a common understanding could be used to inform the development, updating and quality assurance of pre-service teacher education programmes and continuous professional learning on data literacy. The conceptual framework of data literacy by Mandinach and Gummer (2016[155]) may provide a useful starting point for defining the concept and underlying skills (see Table 4).

Table 4. Example of a conceptual framework of data literacy – selection of skills

Identify problems/frame questions	Use data	Transform data into information	Transform data into decisions	Evaluate outcomes
• Articulate a problem of practice about a student/aspect of instruction • Understand the context at the student and school level • Involve other stakeholders in the process • Understand student privacy and confidentiality	• Identify possible sources of data • Understand how to generate data • Understand elements of data accuracy, appropriateness and completeness • Find, locate, access and retrieve data • Manage, organise and prioritise data	• Generate hypothetical connections to instruction • Understand how to interpret data • Understand and use data displays and representations • Articulate inferences and conclusions • Synthesise diverse data	• Determine the next instructional steps • Monitor student performance • Diagnose what students need • Make instructional adjustments • Understand the context for the decision	• Re-examine the original question or problem • Compare performance pre- and post-decision • Monitor changes in classroom practices • Monitor student changes in performance • Consider the need for iterative decision cycles

Source: Mandinach and Gummer (2016[155]), *What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions*, https://doi.org/10.1016/j.tate.2016.07.011.

Ensuring data literacy skills development in pre-service teacher education

Teachers and school leaders must be trained to have a strong foundation of data literacy skills, to ensure not only an understanding of data collection and statistical techniques but also how that data should be used to inform teaching and student learning (Henderson and Corry, 2021[159]). Pre-service teacher education programmes should ensure that aspiring teachers are trained in and supported in developing data literacy skills. This calls for actively integrating data literacy skills and pedagogy into pre-service programmes (Mandinach and Gummer, 2012[164]; Hunter-Thomson, 2022[165]).

The introduction of the standardised student assessments may provide an opportunity to explore whether data literacy skills are sufficiently covered and developed in pre-service teacher education in the Flemish Community. In fact, the Flemish Community is arguably well-positioned to ensure adequate integration of data literacy in its teacher education programmes. The central role of universities in the reform (through the Steunpunt) could support prompt updating of teacher education programmes in these institutions where needed, in line with a common understanding of data literacy and underlying skills.

Emphasising collaborative professional learning in the effective use of data for informing teaching

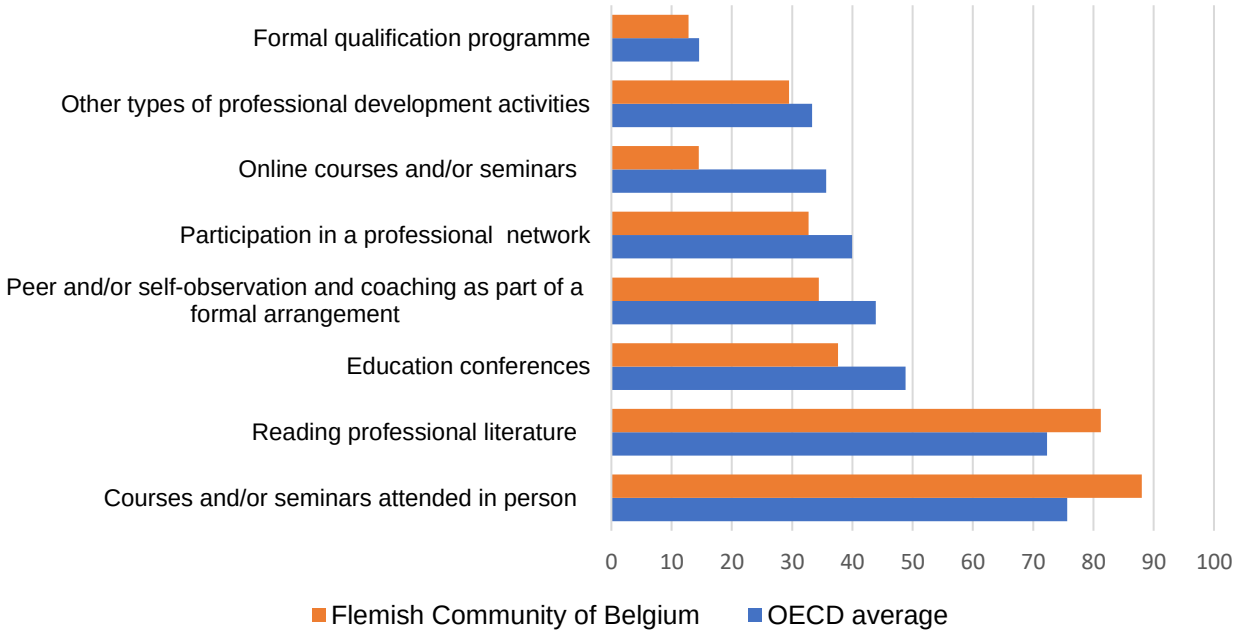
A recent OECD report (OECD, 2021[4]) on the Flemish continuous professional learning system underlines the need for a system-level commitment to strengthen teachers' continuing professional learning, with the potential to build on pockets of excellence within the system. The report pointed to the various Flemish and international studies that have identified a range of challenges facing the Flemish teachers' continuous professional learning system (OECD, 2019[121]; Flemish Department of Education and Training, 2021[166]). It pointed, among others, to the following key challenges: limited engagement in professional learning and the need for enhancing evidence-informed practice, school leadership, collaborative professional learning practices within and between schools, and data literacy. These findings were corroborated by the OECD's

interviews with stakeholders, with several stakeholders specifically expressing their concerns about teachers' and school leaders' data literacy skills.

TALIS 2018, for example, shows that teachers in the Flemish Community devote considerably less time to their professional learning than their peers in other OECD countries. Also, teacher participation in more effective forms of professional learning (e.g. collaborative professional learning, coaching and teacher networks, observation visits to other schools, and in their own school) remains limited relative to more traditional activities, such as one-off courses or seminars, which are known to be less impactful (OECD, 2019[121]) (see Figure 14 and Figure 15). But research evidence shows the importance of teachers and school leaders engaging in collaborative inquiry to improve their data literacy skills (Piro and Hutchinson, 2014[167]; Schildkamp and Poortman, 2015[168]). Collaboration is considered an important and valued component in the inquiry process, where educators work together to examine data and seek solutions to a particular problem (Mandinach and Gummer, 2016[155]; OECD, 2019[121]).

Figure 14. Types of professional development undertaken by teachers, 2018

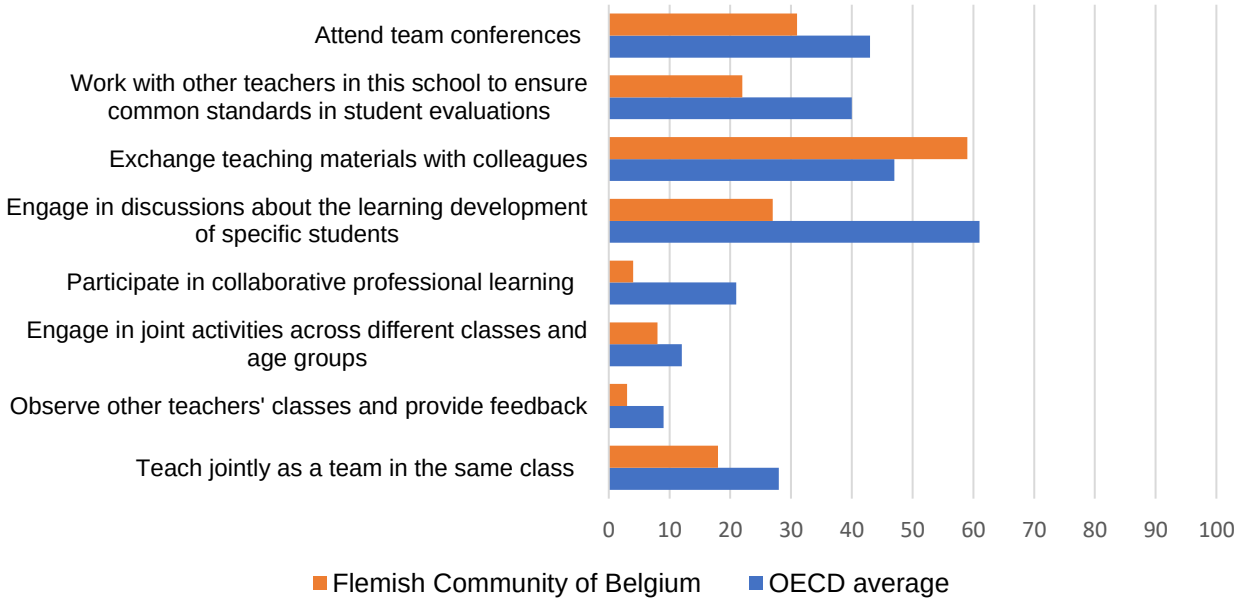
Percentage of lower secondary school teachers who reported having participated in the following professional development activities in the 12 months prior to the survey



Note: The OECD average refers to the average of the 31 OECD countries and economies that participated in TALIS 2018.
Source: OECD (2019[121]), TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners, Table I.5.7., https://doi.org/10.1787/1d0bc92a-en.

Figure 15. Participation in collaborative working and professional learning by teachers, 2018

Percentage of secondary school teachers that reported at least once a month to:



Note: The OECD average refers to the average of the 31 OECD countries and economies that participated in TALIS 2018.
Source: OECD (2020[169]), TALIS 2018 Results (Volume II): Teachers and School Leaders as Valued Professionals, https://doi.org/10.1787/19cf08df-en.

In light of the evidence, the DoET, Steunpunt, PDB and other education stakeholders should (continue to) emphasise collaborative professional learning within and between schools to enhance teachers' and school leaders' data literacy skills. Importantly, the PDB have, in recent years, shifted away from supporting individual teachers to supporting entire schools (OECD, 2021[4]).

Furthermore, while school leaders have an important role to play in setting a school's culture around effective data use, schools' quality assurance and development processes remain variable and, in many cases, underdeveloped (Shewbridge and Köster, 2021[5]). The development of school leaders' data literacy skills and inquisitive mindsets, as well as their skills for developing a collaborative learning culture in their schools geared towards the proactive use of data for improving teaching and student learning are, therefore, important areas of professional learning for school leaders in the Flemish Community.

The need for strategic and sustained investment in the professional learning of teachers' and school leaders' data literacy skills

Informed in part by a recent OECD report (2021[4]), the DoET is aware of the challenges and opportunities of its continuous professional learning system. It has taken several important measures recently, some of which directly or indirectly support teachers' and school leaders' data literacy skills development. As described above, the DoET has been implementing a digitalisation project that includes training teachers (i.e. Digisprong) and the Steunpunt is developing online courses focusing on the new standardised student assessments. The online courses would include support for data literacy (i.e. how to interpret assessment results) and data use (i.e. how to guide learning improvement based on assessment results). The Steunpunt also plans to collaborate with the PDB on further actions to support school developments (Steunpunt, 2022[133]).

The Flemish government has also allocated a yearly investment of EUR 1.5 million for enhancing data literacy skills through the 2023 and 2024 budget of the PDB. However, herein also lies a challenge, given that schools receive a relatively limited annual budget for each teacher's professional learning. In addition, a relatively large share of projects funded by the DoET aimed to support professional learning does not benefit from recurrent funding and are designed as one-off initiatives, which may be too short to achieve or sustain reasonable change for teachers' learning or school policies (OECD, 2021[4]). Considering the professional learning needs of the Flemish education profession and the fragmented continuous professional learning system, a more strategic approach and longer term investment in the development of data literacy skills of the Flemish education profession would seem warranted.

The Flemish government is envisioning further reforms to enhance professional learning within the education profession. A reform of the programme for policy and practice-oriented research (OBPWO) and a reform of the PDB have already been implemented (Flemish Department of Education and Training, 2021[166]). The PDB have (as mentioned) shifted away from supporting individual teachers to supporting entire schools. The reform of the PDB comprises various strands, including supporting teachers' classroom practices (Weyts, 2019[156]). In addition, the Education Policy Note 2019-2024 includes plans to support enhanced collaboration across the PDB, as well as a more regular evaluation of their work.

Apart from striving towards optimal (i.e. cost-effective) use of the limited financial resources available for supporting professional learning and school development efforts more broadly, such collaborations allow pooling expertise, peer learning and sharing good practices across the different PDB (Roy et al., 2021[170]). Therefore, to support the successful implementation of the standardised student assessments, as well as other policies, the PDB should continue expanding and deepening the collaborations between them. The OECD learnt of several good examples from the past to build on, including the "SNPB", a cross-network partnership of four PDB (2006-15) and the "networks of expertise" (2007-15) (OECD, 2021[4]).

The 2021 OECD report also offered a number of options for strengthening the continuous professional learning system that are vital for supporting the successful implementation of the standardised student assessment reform. These include the importance of a co-constructed professional learning strategy for the education profession that could inform a broader vision for the teaching profession in the decades to come, informing all aspects of teacher policy. This strategy should ensure multi-year funding for professional learning and could focus on strengthening teachers' and school leaders' data literacy skills, promote collaborative learning within and across schools, strengthen internal quality assurance processes, and should be demand-driven (i.e. based on the learning needs of teachers and school leaders and school development priorities). Schools' internal quality assurance and development processes (e.g. using the OK Framework), and the external evaluation by the Inspectorate of Education, provide important means for identifying and responding to these needs with the support of the PDB (see Figure 3).

For its development or co-construction of this strategy, the DoET and education stakeholders may look towards the example of the Netherlands, which in 2013 released a comprehensive action plan as a response to a range of challenges similar to those of the Flemish Community of Belgium (i.e. equity concerns, teacher shortages, limited collaborative working and learning within schools, etc.). The action plan, Teacher Agenda 2013-20: The Teacher Makes the Difference, aimed to empower teachers' capacity to respond to these and other challenges, and consisted of seven complementary "action points" (strands of work) that cover the stages of the professional lifecycle (i.e. initial teacher education, recruitment, professional learning and growth) and emphasise the development of a collaborative learning culture (MoECS, 2013[171]). This example could help position the implementation of the assessments as part of a wider policy initiative to promote collaborative working and learning with and between schools, and to attract, retain and develop a high-quality education profession. Importantly, this action plan was the result of a broad stakeholder engagement process involving teachers, school leaders and boards, tertiary education institutions, social partners and others, thereby adding to its quality and relevance while gaining the support of stakeholders for its successful implementation.

Recommendations

- **Continue and further enhance stakeholder engagement and "co-construction" of the reform, with a particular emphasis on engaging school leaders at scale.** Given school leaders' important role in continuous professional learning and leading educational change for school development, their engagement is critical. They could support raising awareness for the tests becoming feedback channels between the schools and the DoET, and, provided they have the capacity, facilitating the effective use of the student assessment data for school development.

  The Steunpunt is a good example of stakeholder engagement and co-construction that should be continued, and possibly be expanded on as the reform advances from a focus on assessment design into one more focused on the administration, analysis, reporting and actual use of the assessments results in schools.
- **The DoET should urgently consider developing and operationalising a comprehensive communication strategy that promotes a cohesive and positive narrative around the standardised student assessment reform.** Central to this is clarifying the main purpose(s) of the test. The limited communication on the reform and the realisation that the large-scale pilot (the calibration) of the assessments is scheduled for May 2023 and the first administration a year later argue for urgency in the operationalisation of the strategy.
  - **The narrative should explain "why" and "how" the assessments – and other policy measures – can support the realisation of the overarching goal of improving the quality of education.** It should aim to concretise how the assessments can be used to improve the quality of education, for example, by focusing on answering questions such as "How will the data be used by schools to support student learning?"; "How will the Inspectorate of Educationuse these data?"; "How does the test complement and add to already existing student assessments?".

    To help clarify the rationale (i.e. the "why") of the reform, part of the communication could highlight the fact that the Flemish Community doesn't have central exams at the end of compulsory education, nor census-based standardised student assessments in support of system-level monitoring and improvement efforts at different levels of the system, making it an outlier among OECD countries. Further, the adaptive nature of the assessments, measurement of learning gains and planned use of contextualised school results are complex issues that should be well explained in an easy-to-understand manner.
  - Recognising the equity challenges of the Flemish school system, **the overarching reform goal of improving the quality of education could be expanded to improving quality and equity in education.** The standardised student assessments aim to introduce reliable and objective student performance results across all schools in the Flemish Community, which would allow identifying and supporting those schools that need it most.
  - **The DoET should work with the education profession, PDB, the Inspectorate of Education and other partners to develop a coherent narrative that is communicated by all.** The DoET and education stakeholders should collaborate to ensure consistent messaging; optimise reach; and use multiple media channels based on reach, frequency and credibility.
  - **The communication strategy should celebrate successes, as these may help raise awareness, support and, ultimately, favour the successful implementation of the reform.**
  - **It will be important to communicate that in the test administration, things may not initially go perfectly, noting that this is to be expected, and highlight the importance of seeing possible implementation challenges as a "learning experience" from which lessons learnt will be taken forward in subsequent rounds of testing,** pointing to the process evaluations that are already planned for the piloting and first year(s) of implementation.
  - **The DoET should develop additional two-way communication channels that create opportunities for a wide range of stakeholders, including teachers and students, to provide feedback and comments beyond the introduction of the assessments.** The DoET should facilitate independent monitoring of the awareness, "readiness" and support for the assessments among teachers, school leaders, students and their parents, which may inform further targeting of communication.
- **The Steunpunt should – as planned – develop official guidelines that clearly describe the roles, responsibilities and expectations of everyone involved in the test administration** (i.e. teachers, school leaders, ICT co-ordinators, students).

Although not urgent, **the DoET should consider how to best govern and deliver the standardised student assessments in the medium- to longer term.** Whether to (in due time) opt for a dedicated autonomous agency or establish this within the government structure is to be carefully considered against the possible advantages and disadvantages of these options. It also seems vital to build on the expertise available in the Steunpunt.

Whatever construct is decided upon, it should have strong governance and oversight so that reform objectives are achieved; policy and practice align; and risks are identified, shared, monitored and managed across all relevant bodies.

- Recognising the identified professional learning needs of the Flemish education profession and the fragmented continuous professional learning system, **a more strategic approach and longer term investment in the development of the data literacy skills of the Flemish education profession would seem warranted.** It is important to continue investing in the development of school leaders' data literacy skills and inquisitive mindset, as well as their skills for developing a collaborative learning culture in their schools geared towards the proactive use of data for quality assurance and improvement processes.
  - **The Flemish standardised student assessment reform may benefit from having the Steunpunt and other education stakeholders define the concept of data literacy and the actual skills involved.** Such a common understanding of data literacy and skills could be used to inform the development, updating and quality assurance of pre-service teacher education programmes and continuous professional learning courses and resources.
  - **Professional learning should also focus on supporting teachers in developing, marking and using self-developed formative and summative assessments.** This could lower the risk of teachers over relying on the new standardised assessments and questioning their own skills and priorities when assessing students' learning needs and performance.
  - **The DoET, the Steunpunt, PDB and education stakeholders should (continue to) emphasise collaborative professional learning within and between schools to enhance teachers' and school leaders' data literacy skills.** To support the successful implementation of the standardised student assessments, as well as other policies, the PDB should also continue expanding and deepening the collaborations between them.
  - An earlier OECD report recommended the co-construction of a professional learning strategy for the education profession that could inform a broader vision for the teaching profession in the decades to come, informing all aspects of teacher policy. **This strategy should ensure multi-year funding for professional learning and focus on strengthening teachers' and school leaders' data literacy skills, promoting collaborative learning within and across schools, and strengthening internal quality assurance and school development processes.** The support should be demand-driven, i.e. based on the learning needs of teachers and school leaders and school development priorities.

4. Developing a comprehensive implementation strategy for the standardised student assessment reform

Building on the strengths and recommendations provided in the previous sections, this concluding section of the report discusses the benefits of developing a comprehensive implementation strategy for the Flemish standardised student assessment reform. The section concludes by proposing and giving concrete guidance for the development of a monitoring, evaluation and research programme to monitor progress in implementing the reform, and support reflection and learning for improving policy and practice at all levels of the system.

A comprehensive implementation strategy

The benefits of developing a comprehensive implementation strategy

The Flemish standardised student assessment reform is a complex, multi-year exercise. Its success will depend on the continued commitment from government and the DoET senior management and strong stakeholder engagement. It will also depend on careful planning, monitoring of progress (see below), risk management, and timely actions to respond to implementation challenges (OECD, 2020[7]; 2020[129]; Viennet and Pont, 2017[9]). The DoET has published a general timeline for the introduction of the standardised student assessments (see Table 2) and has put in motion various strands of work. However, a detailed implementation strategy that brings these different components and underlying policy actions together is lacking.

Research has shown that with increased complexity in education policy making, policy makers need to consider many factors and take appropriate action. Designing and communicating a coherent implementation strategy can help guide the complex web of interactions required for a policy to be realised in schools (OECD, 2020[8]). Therefore, the DoET should consider developing a comprehensive multi-year implementation strategy to help guide the reform and monitor progress against defined objectives and milestones. The strategy should describe the key actions to be taken by implementation partners and other stakeholders. Again, research evidence shows the importance of developing such a strategy with the education profession and other key stakeholders, as it can build consensus, give them a strong sense of ownership and strengthen their confidence in the change process (Schleicher, 2018[12]; Viennet and Pont, 2017[9]). The development of an implementation strategy, as well as other public communications, can be understood as an instrument for policy making and communication, trust-building and, ultimately, as a means to strengthen the success of the reform (OECD, 2019[172]; Brezzi et al., 2021[173]).

In the strategy, it is essential to consider the time, equipment and facilities, and financial resources necessary to support the implementation and sustainability of the reform (OECD, 2020[8]). Ensuring multi-year funding will be particularly important for supporting teachers' and school leaders' development of data literacy skills and for strengthening internal quality assurance and school development processes (see Section 3), as well as for the proposed monitoring, evaluation and research programme (see below).

It is also important to recognise the roles and responsibilities different stakeholders, like the PDB, the Steunpunt and the Inspectorate of Education, play in the implementation of different policy actions. The explicit recognition of these implementation partners and their awareness of each other's roles and responsibilities could benefit the successful implementation of the reform (Viennet and Pont, 2017[9]).

The fact that large-scale piloting is taking place in May 2023 and the first administration the following year argues for urgency in developing the standardised student assessment reform implementation strategy. This report has identified several strengths of the design and implementation of the standardised student assessment reform that should be maintained and built upon, as well as several areas for improvement for which concrete recommendations are offered. These could form the basis for developing a comprehensive implementation strategy to guide the reform through the various stages of implementation, with the necessary sequencing and timing of policy actions, outputs, milestones and objectives to optimise the capacity of all involved and avoid overload and/or partial implementation.

As mentioned, the monitoring of progress and timely identification of implementation challenges (i.e. risk management) is an essential component of an implementation strategy. This section offers concrete recommendations for developing a comprehensive monitoring, evaluation and research programme.

Ensuring ongoing monitoring, evaluation and research to support successful implementation of the reform

The draft decree on the standardised student assessments notes the importance of monitoring and evaluating the implementation of the standardised student assessment reform (Flemish Government, 2023[53]). It states that in 2023 (as part of the calibration), a process evaluation will be carried out on a representative sample of schools. The process of the first student assessment administration and feedback will be evaluated so that the process of testing and feedback proceeds as well as possible in future years. The first assessments will be administered across all Flemish schools in April-May 2024. The evaluation will take place immediately afterwards and the lessons learnt will be taken forward in the following assessment cycles. An overall evaluation of the effects of the assessments on the reform purposes and any other effects is planned at a later stage. This subsection provides recommendations for further monitoring, evaluation and research to support the successful implementation of the reform.

Determining and evaluating the validity of the standardised student assessments

Experts agree that validity is the most important concept in the field of educational assessment, but there is much less consensus as to how validity ought to be defined (Newton and Baird, 2016[174]). The closest there is to a consensus definition is arguably provided by the Standards for Educational and Psychological Testing, which noted that "*validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of assessments"* (AERA, APA and NCME, 2014, p. 11[65]). It is clear from this definition that it will be necessary to validate the assessments for the intended purposes.

The debate about the definition of validity is partly centred on its scope. Whether it ought to relate only to measurement quality and, as such, is a fixed property of a test; or be seen as an overarching evaluative concept, addressing the overall defensibility or acceptability of the policy which endorses the use of test scores for a particular purpose. The latter definition extends validity to encompass both unintended and intended consequences arising from test score use (Newton and Shaw, 2015[175]). Evaluating unintended consequences requires consideration of unintended uses of test results that may emerge (see below).

It is good practice to publicly report the measurement quality of standardised assessments annually. Such transparency may help build confidence, awareness and understanding of the standardised student assessment reform and the objectives it has set out to achieve. As planned, the Steunpunt is in a good position to undertake and produce such annual test validation reports, having both the necessary expertise and access to the data. In addition, confidence in the validation exercises could be built by encouraging other experts in educational assessment from the academic and other communities to evaluate the measurement quality of the assessments (i.e. external validation of the validity and reliability of the assessment).

Monitoring and evaluating the reform

Interviews with education stakeholders corroborated earlier findings (Shewbridge and Köster, 2021[5]) that education stakeholders in the Flemish Community were generally positive about using the standardised assessment data for system-level monitoring. Several stakeholders noted the importance of ensuring that the test results do not come to dominate – and narrow the view and discussions of – education quality in the Flemish Community. They also noted that the findings should serve the purpose of "learning" and supporting development and improvement in policy and practice. The DoET and all other education stakeholders seemed to share these views.

Furthermore, it will be important for the DoET and education stakeholders to reflect on how assessment results are going to be interpreted, including the variables that may affect changes over time and the interpretation of these data at the system level. Explaining the methodology, rationale and expectations ahead of the administration of the assessments may increase support and credibility. Such a discussion would help interpret changes in outcomes over time. This is vital as there is a distinction to be drawn between students' performance on a student assessment and their underlying attainment (i.e. what they know and/or can do). There are many reasons why performance and attainment might diverge. For example, anxiety and student motivation might be influential factors, which may vary between grades or over time. Student motivation will be explored in the context of the Flemish standardised student assessments by the University of Ghent.

The "sawtooth effect" (Koretz, 2008[176]; 2017[177]) may also be at play. This effect refers to the possibility that improvement in test scores over time is primarily due to successive cohorts becoming better at tackling the kinds of tasks that appear in the test, owing to increasing test familiarity. In other words, while rising student results represent an improvement in cohort performance, they do not represent an improvement in cohort attainment. Often when a testing programme is then reformed, the system-level outcomes drop to the levels observed when the last set of assessments were introduced. The sawtooth effect may also suggest an apparent improvement in student assessment outcomes in the first two to three years of assessments, followed by a levelling off (Cuff, Meadows and Black, 2019[178]). In reality, understanding the reasons for performance improvements is extremely complex (Newton, 2020[179]), with some improvements driven by constructive realignment and others by the kinds of test preparation strategies that are likely to be frowned upon. Any possible unintended consequences need to be monitored and considered in interpreting the system-level outcomes.

In addition, international evidence shows the importance and potential benefits of monitoring the social and ethical consequences of the use(s) of the standardised assessments, such as students' motivation and well-being. Currently, such studies are planned to be carried out by the Steunpunt. Funding this research and ensuring it covers all the aforementioned topics, particularly in the early years of the reform, would inform and support a successful implementation.

Building on the importance of evaluation of the standardised student assessment reform already noted in the draft decree, the DoET should consider establishing a multi-year monitoring, evaluation and research programme to monitor progress in implementing the reform, and support reflection and learning for

improving policy and practice at all levels. Engagement with a wide range of education stakeholders should be continued to shape the monitoring, evaluation and research programme and oversee its implementation. The programme should form an integrated part of the implementation strategy.

More concretely, the monitoring, evaluation and research programme could include studies such as:

- Monitoring the skills and readiness of the education profession for the introduction of the assessments and the use of the data to improve their practice and student learning. A detailed examination of teachers' and school leaders' data literacy skills could, for example, greatly inform further professional learning.
- Conduct research into student test taking motivation and their well-being, as fluctuations in motivation over time may impact on assessment results and the interpretation of changes, in particular how they relate to educational standards, i.e. attainment targets.
- Monitoring of unintended consequences, such as narrowing of the curriculum, test preparation and malpractice (see, for example, Meadows and Black (2018[180])).
- Surveys of public confidence and trust in the assessments.
- Monitoring of a potential sawtooth effect in assessment results in the early years of administration (see, for example, Cuff, Meadows and Black (2019[178])).
- Analysis of the relationship between assessment results and grades given by teachers, including by socio-economic status and other characteristics (e.g. age, gender, disability, etc.) to shine light on the reliability of results (see, for example, Lee and Newton (2021[181])). Gaps between test results and teacher grades may raise further questions for research, such as the possibility of construct under-representation by the test or bias in grading.
- Alignment research to evaluate the extent of the connection between the assessments, standards and teaching practices (Martone and Sireci, 2009[182]).
- Analysis of the relationship between changes in the Flemish assessment results and those of international student assessments to aid interpretation of changes in students' performance.

These and other possible studies, several of which could serve as macro-validation of the assessments (Newton, 2016[183]), could complement the micro-validation of elements of the assessment process.

The Steunpunt shared several of its plans for further research, many of which align with these points. One of these studies focused on examining the key characteristics of well-performing or effective schools, their context and other factors that influence student results. During the visit to the Flemish Community, the OECD and Steunpunt researchers explored the idea of incorporating the OK Framework in such a study, given its importance as a conceptualisation of "a quality school" in the Flemish context. The deliberate pursuit of such key policy/research questions and policy coherence will be vital for facilitating "learning" and informing the improvement of policy and practice. In addition, following the examples of countries like Ireland and the Netherlands (Dutch Inspectorate of Education, 2022[184]), the DoET and the Inspectorate of Education should consider using the monitoring, evaluation and research programme and its findings to enrich the Inspectorate of Education's annual report, the *Education Mirror* (Dutch Inspectorate of Education, 2021[26]).

In sum, investing in a comprehensive monitoring, evaluation and research programme could have many positive outcomes, including the ability to: respond quickly to unintended negative consequences; appropriately interpret changes in test results; build public confidence in the reform by demonstrating a commitment to evidence-based policy making; and identify and share successes and good practices to support learning, as well as those issues that call for policy responses to further empower schools in their improvement efforts.

Recommendations

- **The DoET should consider developing a comprehensive multi-year implementation strategy to help guide the reform and monitor progress.** The fact that large-scale piloting of the test is taking place in May 2023 and the first administration is scheduled for the following school year argues for urgency in developing the implementation strategy. Doing this with the profession and other key education stakeholders could help build consensus and ownership, and strengthen confidence in the reform.
  - This report has identified several strengths of the design and implementation of the standardised student assessment reform that should be maintained and built upon. It has also identified several areas for improvement for which concrete recommendations are offered. **The identified strengths and recommendations for improvement could form the basis for a comprehensive multi-year implementation strategy**, with the necessary sequencing and timing of policy actions, outputs and outcomes to optimise the capacity of all involved.
- **The DoET should consider establishing a multi-year monitoring, evaluation and research programme to monitor progress in implementing the reform and support reflection and learning for improving policy and practice at all levels.** The research could focus on:
  - **The validity of the assessments: The Steunpunt is – as planned – in an excellent position to produce annual test validation reports, having both the necessary expertise and access to the data.** That said, confidence in the validation exercises would be reinforced by encouraging other experts to evaluate the quality of the assessments (i.e. external validation of the validity and reliability of the assessment).
  - **The social and ethical consequences of the use(s) of the standardised assessments, including unintended consequences,** for example, by examining:
    - The skills and readiness of the education profession for introducing the assessments and using the data to improve their practice and student learning. A detailed examination of teachers' and school leaders' data literacy skills could, for example, greatly inform further professional learning.
    - Students' motivation for taking the assessments and their well-being, as fluctuations over time may have an impact on test results.
    - Unintended consequences such as narrowing of the curriculum, test preparation and malpractice.
    - The public's confidence and trust in the assessments.
    - Potential "sawtooth effect" in test results in the early years of administration.
    - The equity dimension of the standardised student assessments.
  - **The DoET and education stakeholders must reflect on how the assessment results are going to be interpreted, including the variables that may affect changes over time and the interpretation of these data at the system level.** Explaining the methodology, rationale and expectations ahead of the administration of the assessments may increase support and credibility**.**
  - **The deliberate pursuit of key policy/research questions and considering policy coherence will be vital for facilitating "learning" and informing the improvement of policy and practice at all levels of the system**.
- Following the examples of countries like Ireland and the Netherlands, **the DoET and the Inspectorate of Education could consider using the research programme and its findings to enrich the Inspectorate of Education's annual report, the *Education Mirror***.

Implementing policies: Supporting change in education

This document was prepared by the Implementing Education Policies Project team, in close collaboration with the Strategic Education Governance Project team at the OECD.

The OECD project Implementing Policies: Supporting Effective Change in Education offers peer-learning and tailored support for countries and jurisdictions to help them achieve success in the implementation of their policies and reforms in school education. The tailored support consists of three complementary strands of work that target countries' and jurisdictions' needs: policy and implementation assessment, strategic advice, and implementation seminars.

For more information

Contact: Marco Kools, project manager and analyst**,** marco.kools@oecd.org

Website: OECD Implementing Education Policies

References

ACARA (2021), "My school", web page, https://www.myschool.edu.au (accessed on 12 September 2022). [80]

ACARA (2015), *NAPLAN Online Research and Development. Report 1: Device Effect Study – Literature Review. Report 2: Device Effect Study – Field Trial*, Australian Curriculum, Assessment and Reporting Authority, https://nap.edu.au/docs/default-source/default-document-library/naplan-online-device-effect-study.pdf. [110]

ACARA (2014), *Tailored Test Design Study 2013: Summary Research Report*, Australian Curriculum, Assessment and Reporting Authority, https://nap.edu.au/docs/default-source/default-document-library/tailored_test_design_study_2013_summary_research_report.pdf. [109]

ACARA (n.d.), *ACARA website*, https://www.acara.edu.au/about-us (accessed on 14 March 2023). [149]

ACARA (n.d.), "Research and development", web page, https://www.nap.edu.au/naplan/understanding-online-assessment/research-and-development#SEM (accessed on 20 September 2022). [112]

ACER and UNESCO Institute for Statistics (2017), *Principles of Good Practice in Learning Assessment*, ACER, https://gaml.uis.unesco.org/wp-content/uploads/sites/2/2018/10/4.1.1_05_Manual-of-Good-Practices-in-Learning-Assessment.pdf. [142]

AERA, APA and NCME (2014), *Standards for Educational and Psychological Testing*, American Educational Research Association, American Psychological Association and National Council on Measurement in Education. [65]

AHOVOKS (n.d.), *Primary education educational goals*, https://onderwijsdoelen.be/resultaten?onderwijsstructuur=LO&filters=onderwijsniveau%255B0%255D%255Bid%255D%3Df7dcdedc9e9c97a653c7dba05896ef57a333480b%26onderwijsniveau%255B0%255D%255Btitel%255D%3DBasisonderwijs%26onderwijsniveau%255B0%255D%255Bwaarde%255D (accessed on 13 October 2022). [21]

Airola, D. and K. Dunn (2011), *Oregon DATA Project Final Evaluation Report*. [152]

Alderson, J. and D. Wall (1993), "Does washback exist?", *Applied Linguistics*, Vol. 14/2, pp. 115-129, https://doi.org/10.1093/applin/14.2.115. [95]

Althaus, C. et al. (2021), "Research and education in public sector practice: A systems approach to understanding policy impact", *Policy Design and Practice*, pp. 309-322, https://doi.org/10.1080/25741292.2021.1977478. [138]

Amrein-Beardsley, A. and J. Holloway (2017), "Value-added models for teacher evaluation and accountability: Commonsense assumptions", *Educational Policy*, Vol. 33/3, pp. 516-542, https://doi.org/10.1177/0895904817719519. [86]

Ansell, C., E. Sørensen and J. Torfing (2017), "Improving policy implementation through collaborative policymaking", *Policy and Politics*, Vol. 45/3, pp. 467-486, https://doi.org/10.1332/030557317X14972799760260. [139]

Armenakis, A. et al. (2007), "Organizational change recipients' beliefs scale: Development of an assessment instrument", *The Journal of Applied Behavioral Science*, Vol. 43/4, pp. 481-505, https://doi.org/10.1177/0021886307303654. [136]

Australian Department of Education and Training (2019), *Departmental Report: Review of the Australian Curriculum, Assessment and Reporting Authority*, Australian Government, https://files.eric.ed.gov/fulltext/ED592556.pdf. [145]

Azevedo, J. et al. (2022), "The unequal impacts of COVID-19 on student learning", in Reimers, F. (ed.), *Primary and Secondary Education During Covid-19*, Springer, Cham, https://doi.org/10.1007/978-3-030-81500-4_16. [32]

Behrstock-Sherratt, E., K. Biggers and J. Fetters (2012), "Lessons learned on communication and engagement for educator evaluation: Colorado case study", *Policy to Practice Brief*, National Comprehensive Center for Teacher Quality, https://files.eric.ed.gov/fulltext/ED543667.pdf. [144]

Biggs, J. (1996), "Enhancing teaching through constructive alignment", *Higher Education*, Vol. 32/3, pp. 347-364, https://doi.org/10.1007/BF00138871. [94]

Binkley, M. et al. (2011), "Defining twenty-first century skills", in *Assessment and Teaching of 21st Century Skills*, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-007-2324-5_2. [120]

Blomkamp, E. (2021), "Systemic design practice for participatory policymaking", *Policy Design and Practice*, Vol. 5/1, pp. 12-31, https://doi.org/10.1080/25741292.2021.1887576. [128]

Brezzi, M. et al. (2021), "An updated OECD framework on drivers of trust in public institutions to meet current and future challenges", *OECD Working Papers on Public Governance*, No. 48, OECD Publishing, Paris, https://doi.org/10.1787/b6c5478c-en. [173]

Brown, M., G. McNamara and J. O'Hara (2016), "Quality and the rise of value-added in education: The case of Ireland", *Policy Futures in Education*, Vol. 14/6, pp. 810-829, https://doi.org/10.1177/1478210316656506. [82]

Burns, T. and F. Köster (2016), *Governing Education in a Complex World*, Educational Research and Innovation, OECD Publishing, Paris, https://doi.org/10.1787/9789264255364-en. [10]

Cabañero-Verzosa, C. (2009), "Using strategic communication to build commitment to reform", in *Building Commitment to Reform through Strategic Communication: The Five Key Decisions*, World Bank. [143]

Cardone, M., P. Falzetti and C. Sacco (2019), "INVALSI data for school system improvement: The value added", Working Paper No. 43/2019. [89]

Care, E. et al. (2018), *Education System Alignment for 21st Century Skills: Focus on Assessment*, Brookings, https://files.eric.ed.gov/fulltext/ED592779.pdf. [93]

Cizek, G. (2005), "High-stakes testing: Contexts, characteristics, critiques, and consequences", in Phelps, R. (ed.), *Defending Standardised Testing*, Lawrence Erlbaum Associates Publishers, https://doi.org/10.4324/9781410612595. [96]

Constitutional Court (2022), *Press Release Judgment 82/2022*, Constitutional Court, https://www.const-court.be/public/n/2022/2022-082n-info.pdf. [47]

Copp, D. (2018), "Teaching to the test: A mixed methods study of instructional change from large-scale testing in Canadian schools", *Assessment in Education: Principles, Policy & Practice*, Vol. 25/5, pp. 468-487, https://doi.org/10.1080/0969594X.2016.1244042. [72]

Cuff, B., M. Meadows and B. Black (2019), "An investigation into the sawtooth effect in secondary school assessments in England", *Assessment in Education: Principles, Policy & Practice*, Vol. 26/3, pp. 321-339, https://doi.org/10.1080/0969594X.2018.1513907. [178]

Darling-Hammond, L. (2020), "Accountability in teacher education", *Action in Teacher Education*, Vol. 42/1, pp. 60-71, https://doi.org/10.1080/01626620.2019.1704464. [74]

Delegates of the Education Providers (2022), *Consultation on Fundamental Educational Reforms as Referred to in Article 11bis (Primary Education Decree) and Article 70 (Codex Secondary Education) on the Preliminary Draft Decree on the Flemish Tests*, Delegates of the Education Providers. [59]

DeMatthews, D., D. Knight and S. Woulfin (2021), ""Show me the data": How schools can appropriately utilize standardized testing for improvement", *TEPSA News*, Vol. 78/5, https://www.tepsa.org/resource/show-me-the-data-how-schools-can-appropriately-utilize-standardized-testing-for-improvement. [29]

Dumont, H., D. Istance and F. Benavides (2011), *The Nature of Learning: Using Research to Inspire Practice*, Educational Research and Innovation, OECD Publishing, Paris, https://doi.org/10.1787/9789264086487-en. [126]

Dutch Inspectorate of Education (2022), *The State of Education in the Netherlands*, Dutch Inspectorate of Education, https://english.onderwijsinspectie.nl/documents/annual-reports/2022/04/28/state-of-education-2022. [184]

Dutch Inspectorate of Education (2021), *Education Mirror 2021 (Onderwijsspiegel 2021)*, Dutch Inspectorate of Education, https://publicaties.vlaanderen.be/view-file/41674. [26]

European Commission (2022), "Plan de Digitalización y Competencias Digitales del Sistema Educativo", web page, https://epale.ec.europa.eu/es/content/plan-de-digitalizacion-y-competencias-digitales-del-sistema-educativo (accessed on 9 March 2022). [163]

Eurydice (2022), "National education systems: Belgium – Flemish Community", https://eurydice.eacea.ec.europa.eu/national-education-systems/belgium-flemish-community/belgium-flemish-community (accessed on 30 August 2022). [19]

Eurydice (2009), *National Testing of Pupils in Europe: Objectives, Organisation and Use of Results*, Eurydice, https://www.eurydice.si/publikacije/National-testing-of-pupils-in-Europe_Objectives,-organisation-and-use-of-results-EN-HI.pdf?_t=1569928954. [64]

Faddar, J. et al. (2020), *Vlaanderen in TIMSS 2019: Wiskunde- en wetenschapsprestaties van het vierde leerjaar in internationaal perspectief en doorheen de tijd (Flanders in Trends in International Mathematics and Science Study 2019)*, Antwerp University, https://data-onderwijs.vlaanderen.be/documenten/bestand.ashx?id=12933. [2]

Faria, A. et al. (2012), *Charting Success: Data Use and Student Achievement in Urban Schools*, Council of the Great City Schools, https://files.eric.ed.gov/fulltext/ED536748.pdf. [153]

Farrell, C. and J. Marsh (2016), "Contributing conditions: A qualitative comparative analysis of teachers' instructional responses to data", *Teaching and Teacher Education*, Vol. 60, pp. 398-412, https://doi.org/10.1016/j.tate.2016.07.010. [160]

Feniger, Y., M. Israeli and S. Yehuda (2015), "The power of numbers: The adoption and consequences of national low-stakes standardised tests in Israel", *Globalisation, Societies and Education*, Vol. 14/2, pp. 183-202, https://doi.org/10.1080/14767724.2015.1010438. [78]

Flemish Department of Education and Training (2022), *Flemish Education in Figures 2020-2021*, Flemish Department of Education and Training, Brussels, https://publicaties.vlaanderen.be/view-file/48569. [14]

Flemish Department of Education and Training (2021), *Teachers' Professional Learning: Country Background Report of the Flemish Community of Belgium for the OECD TPL Study*, Flemish Department of Education and Training, https://www.oecd.org/education/teachers-professional-learning-study/continuing-professional-learning/Flemish-community-belgium-background-report.pdf. [166]

Flemish Department of Education and Training (2019), *The Future Labour Market for Educational Staff in Flanders 2018-2028*, Flemish Department of Education and Training, https://publicaties.vlaanderen.be/view-file/30960 (accessed on 5 September 2022). [42]

Flemish Education Council (2022), "Flemish assessments: Still fundamental bottlenecks: Advice in response to the concept memorandum "Further monitoring education quality through Flemish assessments"". [81]

Flemish Government (2023), *Draft Decree on Flemish Tests in Education – VR 2023 1002 DOC.0133/2QUATER*, Flemish Government. [53]

Flemish Government (2022), *Preliminary Draft Decree on Flemish Assessments in Education*, Flemish Government, https://beslissingenvlaamseregering.vlaanderen.be/document-view/62D1989F8E6C4430A889900C. [48]

Flemish Government (2021), *Concept Note: Subject – Further Monitoring of Education Quality via Flemish Assessments*. [57]

Flemish Ministry of Education and Training (2022), "Flemish Keys. Vlaamse Toetsen.", Flanders is Education & Training, https://onderwijs.vlaanderen.be/nl/beleid/onderwijsregelgeving-in-ontwikkeling/vlaamse-toetsen. [140]

Flemish Ministry of Education and Training (2022), *Mathematics Test in Primary Education 2021*, Flemish Ministry of Education and Training, https://peilingsonderzoek.be/wp-content/uploads/2019/01/WIS_BaO_2021_Brochure-2.pdf. [31]

Flemish Ministry of Education and Training (2022), "Online question time about the Flemish tests – 1 February 2022", web page, https://onderwijs.vlaanderen.be/nl/online-vragenuurtje-over-de-vlaamse-toetsen-1-februari-2022 (accessed on 15 September 2022). [141]

Flemish Ministry of Education and Training (2021), "Additional measures in the fight against the teacher shortage", web page, https://onderwijs.vlaanderen.be/nl/extra-maatregelen-in-de-strijd-tegen-het-lerarentekort (accessed on 5 September 2022). [44]

Flemish Ministry of Education and Training (2021), *Report: First High-Level Forum May 12, 2021*, Flemish Ministry of Education and Training. [132]

Flemish Ministry of Education and Training (2019), *Mathematics Test in the 1st Degree of Secondary Education B-Stream*, Flemish Ministry of Education and Training, https://peilingsonderzoek.be/wp-content/uploads/2018/11/Brochure_wisk_SO1B_2019.pdf. [35]

Flemish Ministry of Education and Training (2018), *Dutch Primary Education Test: Reading-listening-writing*, Flemish Ministry of Education and Training, https://peilingsonderzoek.be/wp-content/uploads/2020/06/Factsheet-NederlandsBaO2018_definitief.pdf. [30]

Flemish Ministry of Education and Training (2018), *Mathematics Test in the 1st Degree of Secondary Education A-Stream*, Flemish Ministry of Education and Training, https://peilingsonderzoek.be/wp-content/uploads/2019/06/P_WiskSO1A_Brochure_lay-out_DEF.pdf. [36]

Flemish Ministry of Education and Training (2018), *OK Education Quality (OK Onderwijs Kwaliteit)*, Flemish Ministry of Education and Training, https://www.onderwijsinspectie.be/nl/het-referentiekader-voor-onderwijskwaliteit-het-ok. [22]

Flemish Ministry of Education and Training (n.d.), "Policy: Education regulations in development – Flemish keys", web page, https://onderwijs.vlaanderen.be/nl/beleid/onderwijsregelgeving-in-ontwikkeling/vlaamse-toetsen (accessed on 31 August 2022). [50]

Flemish Parliament (2021), *Draft Decree on Educational Objectives for the Second and Third Grade of Secondary Education and Various Other Related Measures*, Flemish Parliament, https://docs.vlaamsparlement.be/pfile?id=1639736. [46]

Flemish Parliament (2021), *Policy and Budget Notes: Education and Training – Budget 2022*, Flemish Parliament, https://fin.vlaanderen.be/wp-content/uploads/2021/11/2022_BO_Beleids-en-begrotingstoelichting_Onderwijs-en-Vorming.pdf. [43]

Flemish Parliament (2018), *Draft Decree on Educational Objectives for the First Grade of Secondary Education*, Flemish Parliament, https://docs.vlaamsparlement.be/pfile?id=1430723. [45]

French Ministry of National Education and Youth (2022), *Évaluation Rèperes: Guide des scores, Notice technique*, French Ministry of National Education and Youth, Paris, https://eduscol.education.fr/document/41950/download. [123]

French Ministry of National Education and Youth (n.d.), "Des ressources d'accompagnement à l'issue des évaluations", web page, https://eduscol.education.fr/1491/evaluer-en-milieu-d-annee-au-cp-un-point-d-etape-vers-la-reussite (accessed on 23 January 2022). [122]

Goverment of Flanders (2022), *Flemish Resilience Recovery Plan. Monitoring report measurement moment June 2022. Progress report and cover note. Relanceplan Vlaamse Veerkracht. Monitoringsrapport meetmoment juni 2022. Voortgangsrapport en covernota.*, Goverment of Flanders, https://publicaties.vlaanderen.be/view-file/50335. [108]

Government of British Columbia (n.d.), "Foundation Skills Assessment (FSA)", web page, https://www2.gov.bc.ca/gov/content/education-training/k-12/administration/program-management/assessment/foundation-skills-assessment#brochures (accessed on 25 September 2022). [70]

Government of Flanders (n.d.), "The Digital Leap ("Digisprong")", web page, https://www.vlaanderen.be/en/authorities/flemish-resilience/the-digital-leap-digisprong (accessed on 23 August 2022). [106]

Growling, S. (2022), "The teacher shortage in figures: Already more than twice as many open teaching vacancies as 5 years ago", *Vrt News*, https://www.vrt.be/vrtnws/nl/2022/06/07/het-lerarentekort-in-cijfers (accessed on 3 September 2022). [41]

Gummer, E. and E. Mandinach (2015), "Building a conceptual framework for data literacy", *Teachers College Record*, Vol. 117/4, pp. 1-22, https://doi.org/10.1177/016146811511700401. [161]

Hattie, J. (2008), *Visible Learning. A Synthesis of Over 800 Meta-Analyses Relating to Achievement*, Routledge, https://doi.org/10.4324/9780203887332. [127]

Hemker, B. and R. Feskens (2021), *Pedagogical-psychometric Aspects of Central Assessment in Flanders: Feasibility Study – Lot 1*, Cito, https://data-onderwijs.vlaanderen.be/onderwijsonderzoek/project/1661. [66]

Henderson, J. and M. Corry (2021), "Data literacy training and use for education professionals", *Journal of Research in Innovative Teaching & Learning*, Vol. 14/2, pp. 232-244, https://doi.org/10.1108/JRIT-11-2019-0074. [159]

Hunter-Thomson, K. (2022), "We must integrate data literacy into teacher prep, and it's easier than you think", Edpost, https://www.edpost.com/stories/we-must-integrate-data-literacy-into-teacher-prep-and-its-easier-than-you-think (accessed on 13 March 2023). [165]

Ifenthaler, D. (2022), "Automated essay scoring systems", in *Handbook of Open, Distance and Digital Education*, pp. 1-15, Springer Nature, Singapore, https://doi.org/10.1007/978-981-19-0351-9_59-1. [116]

INVALSI (n.d.), *The INVALSI Tests According to INVALSI*, INVALSI, https://invalsi-areaprove.cineca.it/docs/2018/INVALSI_tests_according_to_INVALSI.pdf. [90]

Jackson, J., R. Adams and R. Turner (2017), "Evidence-based education needs standardised assessment", *The Conversation*, https://theconversation.com/evidence-based-education-needs-standardised-assessment-87937 (accessed on 4 September 2022). [28]

Kimaru, T. (2017), "The impacts of computer adaptive testing from a variety of perspectives", *Journal of Educational Evaluation for Health Proffesionals*, Vol. 14/12, pp. 1-5, https://doi.org/10.3352/jeehp.2017.14.12. [114]

Kin, T. et al. (2015), "Teacher change beliefs: Validating a scale with structural equation modelling", *School Leadership & Management*, Vol. 35/3, pp. 266-299, https://doi.org/10.1080/13632434.2014.962503. [135]

Kolen, M. and R. Brennan (2014), *Test Equating, Scaling, and Linking*, Springer, New York, NY, https://doi.org/10.1007/978-1-4939-0317-7. [104]

Koretz, D. (2017), *The Testing Charade: Pretending to Make Schools Better*, University of Chicago Press. [177]

Koretz, D. (2008), "Measuring up: What educational testing really tells us", *Journal of Educational Measurement*, Vol. 46/3, pp. 363-366, https://doi.org/10.1111/j.1745-3984.2009.00086_1.x. [176]

Lazendic, G., J. Justus and S. Rabinowitz (2018), *NAPLAN Online Automated Scoring Research* [1

Program: Research Report, Australian Curriculum, Assessment and Reporting Authority, https://nap.edu.au/docs/default-source/default-document-library/naplan-online-aes-research-report-final.pdf. [13]

Leaton Gray, S. and N. Kucirkova (2021), "AI and the human in education: Editorial", *London Review of Education*, Vol. 19/1, https://doi.org/10.14324/lre.19.1.10. [98]

Leckie, G. and H. Goldstein (2019), "The importance of adjusting for pupil background in school value-added models: A study of Progress 8 and school accountability in England", *British Educational Research Journal*, Vol. 45/3, pp. 518-537, https://doi.org/10.1002/berj.3511. [87]

Leckie, G. and H. Goldstein (2017), "The evolution of school league tables in England 1992-2016: 'Contextual value-added', 'expected progress' and 'progress 8'", *British Educational Research Journal*, Vol. 43/2, pp. 193-212, https://doi.org/10.1002/berj.3264. [79]

Lee, M. and P. Newton (2021), *Systematic divergence between teacher and test-based assessment: literature review*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/985236/6781_TA_exam_LitRev_FINAL.pdf. [181]

Levy, J. et al. (2019), "Methodological issues in value-added modeling: An international review from 26 countries", *Educational Assessment, Evaluation and Accountability*, Vol. 31/3, pp. 257-287, https://doi.org/10.1007/s11092-019-09303-w. [85]

Lifelong Achievement Group & Martin, A.J. (2015), *Online NAPLAN Testing and Student Motivation: Exploring Adaptive and Fixed Test Formats*, Lifelong Achievement Group, Sydney, https://www.nap.edu.au/_resources/Online_NAPLAN_and_Student_Motivation.PDF. [111]

Looney, J. (2011), "Alignment in complex education systems: Achieving balance and coherence", *OECD Education Working Papers*, No. 64, OECD Publishing, Paris, https://doi.org/10.1787/5kg3vg5lx8r8-en. [92]

Mandinach, E. and E. Gummer (2016), "What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions", *Teaching and Teacher Education*, Vol. 60, pp. 366-376, https://doi.org/10.1016/j.tate.2016.07.011. [155]

Mandinach, E. and E. Gummer (2012), *Navigating the Landscape of Data Literacy: It IS Complex*, WestEd, https://files.eric.ed.gov/fulltext/ED582807.pdf. [164]

Martone, A. and S. Sireci (2009), "Evaluating alignment between curriculum, assessment, and instruction", *Review of Educational Research*, Vol. 79/4, pp. 1332-1361, https://doi.org/10.3102/0034654309341375. [182]

Masters, G. (2019), *Nurturing Wonder and Igniting Passion: Designs for a New School Curriculum, NSW Curriculum Review*, NSW Education Standards Authority, https://nswcurriculumreform.nesa.nsw.edu.au/pdfs/interimreport/chapters/NSW-Curriculum-Review-Interim-Report.pdf. [97]

May, P. (2015), "Implementation failures revisited: Policy regime perspectives", *Public Policy and Administration*, Vol. 30/3-4, pp. 277-299, https://doi.org/10.1177/0952076714561505. [51]

McClelland, T. and J. Cuevas (2020), "A comparison of computer-based testing and paper and [10]

pencil testing in mathematics assessment", *The Online Journal of New Horizons in Education*, Vol. 10/2, pp. 78-89, https://www.tojned.net/journals/tojned/articles/v10i02/v10i02-01.pdf. [113]

McKnight, C. and E. Glennie (2019), *Are You Ready for This? Preparing for School Change by Assessing Readiness*, RTI International, https://doi.org/10.3768/rtipress.2019.pb.0020.1903. [111]

Meadows, M. and B. Black (2018), "Teachers' experience of and attitudes toward activities to maximise qualification results in England", *Oxford Review of Education*, Vol. 44/5, pp. 563-580, https://doi.org/10.1080/03054985.2018.1500355. [180]

MoECS (2013), *Leraren agenda 2013-2020: de leraar maakt het verschil [Teachers Agenda 2013-2020: The Teacher Makes the Difference]*, Ministry of Education, Culture and Science, http://www.delerarenagenda.nl/documenten/publicaties/2015/01/01/lerarenagenda-oktober-2013. [171]

Molenberghs, G. et al. (2022), *Feedback Use of Central Tests. User Research: In-depth Interviews with School Leaders and Teachers*, https://steunpunttoetsen.be/wp-content/uploads/2023/02/Feedbackgebruik-uit-centrale-toetsen.-Gebruikersonderzoek.-Diepte-interviews-bij-schoolleiders-en-leraren.pdf. [54]

Morris, A. (2011), "Student standardised testing: Current practices in OECD countries and a literature review", *OECD Education Working Papers*, No. 65, OECD Publishing, Paris, https://doi.org/10.1787/5kg3rp9qbnr6-en. [63]

Mullis, I. et al. (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. [33]

Mullis, I. et al. (2020), *TIMSS 2019 International Results in Mathematics and Science*, Boston College, TIMSS & PIRLS International Study Center, https://timss2019.org/reports. [34]

Munoz-Chereau, B., A. Anwandter and S. Thomas (2019), "Value-added indicators for a fairer Chilean school accountability system: A pending subject", *Journal of Education Policy*, Vol. 35/5, pp. 602-622, https://doi.org/10.1080/02680939.2019.1598584. [83]

National Council on Measurement in Education (2019), "Adaptive test", https://www.ncme.org/search?s=Computerised%20adaptive%20test&cs=null&l=1&expanded-categories=undefined (accessed on 3 March 2023). [49]

Newton, P. (2020), *What is the Sawtooth Effect? The Nature and Management of Impacts from Syllabus, Assessment, and Curriculum Transitions in England*, Ofqual, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/936339/What_is_the_Sawtooth_Effect.pdf. [179]

Newton, P. (2017), "There is more to educational measurement than measuring: The importance of embracing purpose pluralism", *Educational Measurement: Issues and Practice*, Vol. 36/2, pp. 5-15, https://doi.org/10.1111/emip.12146. [61]

Newton, P. (2016), "Macro- and micro-validation: Beyond the 'five sources' framework for classifying validation evidence and analysis", *Practical Assessment, Research, and Evaluation*, Vol. 21/12, https://doi.org/10.7275/f75k-1y75. [183]

Newton, P. and J. Baird (2016), "The great validity debate", *Assessment in Education: Principles,* [17]

Policy & Practice, Vol. 23/2, pp. 173-177, https://doi.org/10.1080/0969594x.2016.1172871. [174]

Newton, P. and S. Shaw (2015), "Disagreement over the best way to use the word 'validity' and options for reaching consensus", *Assessment in Education: Principles, Policy & Practice*, Vol. 23/2, pp. 178-197, https://doi.org/10.1080/0969594x.2015.1037241. [175]

Nogueira, A. and R. Schmidt (2022), "Participatory policy design: Igniting systems change through prototyping", *Policy Design and Practice*, pp. 32-50, https://doi.org/10.1080/25741292.2021.1888399. [130]

Nusche, D. et al. (2015), *OECD Reviews of School Resources: Flemish Community of Belgium 2015*, OECD Reviews of School Resources, OECD Publishing, Paris, https://doi.org/10.1787/9789264247598-en. [15]

OECD (2022), *Building Trust to Reinforce Democracy: Main Findings from the 2021 OECD Survey on Drivers of Trust in Public Institutions*, Building Trust in Public Institutions, OECD Publishing, Paris, https://doi.org/10.1787/b407f99c-en. [146]

OECD (2021), "Teachers' professional learning study: Diagnostic report for the Flemish Community of Belgium", *OECD Education Policy Perspectives*, No. 31, OECD Publishing, Paris, https://doi.org/10.1787/5cc2d673-en. [4]

OECD (2020), "An implementation framework for effective change in schools", *OECD Education Policy Perspectives*, No. 9, OECD Publishing, Paris, https://doi.org/10.1787/4fd4113f-en. [8]

OECD (2020), *Innovative Citizen Participation and New Democratic Institutions: Catching the Deliberative Wave*, OECD Publishing, Paris, https://doi.org/10.1787/339306da-en. [7]

OECD (2020), *PISA 2018 Results (Volume V): Effective Policies, Successful Schools*, PISA, OECD Publishing, Paris, https://dx.doi.org/10.1787/ca768d40-en. [17]

OECD (2020), *Policy Framework on Sound Public Governance: Baseline Features of Governments that Work Well*, OECD Publishing, Paris, https://doi.org/10.1787/c03e01b3-en. [129]

OECD (2020), *TALIS 2018 Results (Volume II): Teachers and School Leaders as Valued Professionals*, TALIS, OECD Publishing, Paris, https://doi.org/10.1787/19cf08df-en. [169]

OECD (2019), *Communicating Open Government: A How to Guide*, OECD, Paris, https://www.opengovpartnership.org/wp-content/uploads/2019/05/OECD_OGP_CommsGuide_22052019.pdf. [172]

OECD (2019), *PISA 2018 Results (Volume I): What Students Know and Can Do*, PISA, OECD Publishing, Paris, https://dx.doi.org/10.1787/5f07c754-en. [1]

OECD (2019), *PISA 2018 Results (Volume II): Where All Students Can Succeed*, PISA, OECD Publishing, Paris, https://dx.doi.org/10.1787/b5fd1b8f-en. [37]

OECD (2019), *TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners*, TALIS, OECD Publishing, Paris, https://doi.org/10.1787/1d0bc92a-en. [121]

OECD (2018), *Developing Schools as Learning Organisations in Wales*, OECD Publishing, Paris, https://doi.org/10.1787/9789264307193-en. [56]

OECD (2018), *Education at a Glance 2018: OECD Indicators*, OECD Publishing, Paris, [16]

https://doi.org/10.1787/eag-2018-en.

OECD (2016), *PISA 2015 Results (Volume I): Excellence and Equity in Education*, PISA, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264266490-en. [38]

OECD (2015), *Education at a Glance 2015: OECD Indicators*, OECD Publishing, Paris, https://doi.org/10.1787/eag-2015-en. [25]

OECD (2015), "School education in the Flemish Community of Belgium", in *OECD Reviews of School Resources: Flemish Community of Belgium 2015*, OECD Publishing, Paris, https://doi.org/10.1787/9789264247598-5-en. [20]

OECD (2014), *PISA 2012 Results: What Students Know and Can Do (Volume I, Revised Edition, February 2014): Student Performance in Mathematics, Reading and Science*, PISA, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264208780-en. [39]

OECD (2013), *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*, OECD Reviews of Evaluation and Assessment in Education, OECD Publishing, Paris, https://doi.org/10.1787/9789264190658-en. [6]

OECD (2010), *PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I)*, PISA, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264091450-en. [40]

OECD (2008), *Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added of Schools*, OECD Publishing, Paris, https://doi.org/10.1787/9789264050259-en. [60]

OECD (forthcoming), *Enabling Factors for Effective and Equitable Digital Education: State of Play and Promising Policies (working title)*, OECD Publishing, Paris, forthcoming. [162]

Ofqual (2020), *Online and On-screen Assessment in High Stakes, Sessional Qualifications: A Review of the Barriers to Greater Adoption and How These Might Be Overcome*, Ofqual, Coventry, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/943382/Barriers_to_online_111220.pdf. [102]

OK Education Quality (2016), *The Reference Framework for Quality in Education: Quality Expectations and Quality Images*, https://www.onderwijsinspectie.be/sites/default/files/2022-06/OK_magazine_eng.pdf. [24]

Opposs, D. et al. (2020), "Governance structure and standard setting in educational assessment", *Assessment in Education: Principles, Policy & Practice*, Vol. 27/2, pp. 192-214, https://doi.org/10.1080/0969594x.2020.1730766. [150]

Osborn, M. (2006), "Changing the context of teachers' work and professional development: A European perspective", *International Journal of Educational Research*, Vol. 45/4-5, pp. 242-253, https://doi.org/10.1016/j.ijer.2007.02.008. [157]

Pasquini, R. and C. DeLuca (2022), "Standardized testing vs teachers' classroom assessment practices", *Canadian Teacher Magazine*, https://canadianteachermagazine.com/2022/09/26/standardized-testing-vs-teachers-classroom-assessment-practices. [158]

Penninckx, M. (2019), "Quality education without central examinations: The case of Flanders", [58]

Management in Education, Vol. 34/1, pp. 40-43, https://doi.org/10.1177/0892020619881045.

Penninckx, M. et al. (2017), *Zicht op leerwinst: Scenario's voor gestandaardiseerd toetsen (Scenarios for Standardised Tests)*, Acco, https://medialibrary.uantwerpen.be/oldcontent/container20507/files/Bijlage%20bij%20het%20boek%20-%20Landenstudies.pdf. [62]

Perelman, L. (2020), "The BABEL generator and E-rater: 21st century writing constructs and automated essay scoring (AES)", *Journal of Writing Assessment*, Vol. 13/1, https://escholarship.org/uc/item/263565cq#main. [118]

Perelman, L. (2017), *Automated Essay Scoring and NAPLAN: A Summary Report*, https://aeunt.org.au/uploads/assets/uploads/automated_essay_scoring_and_naplan.pdf. [117]

Perryman, J. et al. (2011), "Life in the pressure cooker – School league tables and English and mathematics teachers' responses to accountability in a results-driven rea", *British Journal of Educational Studies*, Vol. 59/2, pp. 179-195, https://doi.org/10.1080/00071005.2011.578568. [75]

Piro, J. and C. Hutchinson (2014), "Using a data chat to teach instructional interventions: Student perceptions of data literacy in an assessment course", *The New Educator*, Vol. 10/2, pp. 95-111, https://doi.org/10.1080/1547688X.2014.898479. [167]

Poortman, C. and K. Schildkamp (2016), "Solving student achievement problems with a data use intervention for teachers", *Teaching and Teacher Education*, Vol. 60, pp. 425-433, https://doi.org/10.1016/j.tate.2016.06.010. [154]

Rafferty, A. and A. Minbashian (2018), "Cognitive beliefs and positive emotions about change: Relationships with employee change readiness and change-supportive behaviors", *Human Relations*, Vol. 72/10, https://doi.org/10.1177/0018726718809154. [134]

Raudienė, I., L. Kaminskienė and L. Cheng (2022), "The education and assessment system in Lithuania", *Assessment in Education: Principles, Policy & Practice*, Vol. 29/3, pp. 383-394, https://doi.org/10.1080/0969594x.2022.2103516. [77]

Reardon, S. et al. (2018), "The relationship between test item format and gender achievement gaps on math and ELA tests in fourth and eighth grades", *Educational Researcher*, Vol. 47/5, pp. 284-294, https://doi.org/10.3102/0013189x18762105. [119]

Reedy, D. (2019), *Independent Review of the Scottish National Standardised Assessments at Primary 1*, https://www.gov.scot/binaries/content/documents/govscot/publications/progress-report/2019/06/scottish-national-standardised-assessments-review-2019/documents/independent-review-of-the-scottish-national-standardised-assessments-at-primary-1/independent-rev. [148]

Revelle, W. and D. Condon (2019), "Reliability from α to ω: A tutorial", *Psychological Assessment*, Vol. 31/12, pp. 1395-1411, https://doi.org/10.1037/pas0000754. [68]

Ro, J. (2019), "Learning to teach in the era of test-based accountability: A review of research", *Professional Development in Education*, Vol. 45/1, pp. 87-101, https://doi.org/10.1080/19415257.2018.1514525. [73]

Rouw, R. et al. (2016), "United in diversity: A complexity perspective on the role of attainment [18]

targets in quality assurance in Flanders", *OECD Education Working Papers*, No. 139, OECD Publsihing, Paris, https://dx.doi.org/10.1787/5jlrb8ftvqs1-en.

Roy, S. et al. (2021), "Teachers' professional learning study: Diagnostic report for Wales", *OECD Education Working Papers*, No. 33, OECD Publishing, Paris, https://doi.org/10.1787/caf912c7-en. [170]

Sandene, B. et al. (2005), *Online Assessment in Mathematics and Writing: Reports from the NAEP Technology-Based Assessment Project*, Research and Development Series, Department of Education, National Center for Education Statistics, US Government Printing Office, Washington, DC, https://files.eric.ed.gov/fulltext/ED485780.pdf. [105]

Schildkamp, K. and M. Lai (2012), "Conclusions and a data use framework", in *Data-based Decision Making in Education*, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-007-4816-3_10. [151]

Schildkamp, K. and C. Poortman (2015), "Factors influencing the functioning of data teams", *Teachers College Record*, Vol. 117/4, pp. 1-42, https://doi.org/10.1177/016146811511700403. [168]

Schiltz, F. et al. (2018), "The added value of more accurate predictions for school rankings", *Economics of Education Review*, Vol. 67, pp. 207-215, https://doi.org/10.1016/j.econedurev.2018.10.011. [84]

Schleicher, A. (2018), *World Class: How to Build a 21st-Century School System*, Strong Performers and Successful Reformers in Education, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264300002-en. [12]

SGI Europe (2021), *SGIs for the EU Recovery: The Contribution of SGI Providers to the Resilience and Recovery Plans*, SGI Europe, Brussels, https://sgieurope.org/files/SGI%20Europe%20-%20Contribution%20to%20the%20COVID-19%20Recovery.pdf. [107]

Shannon, G. and P. Bylsma (2007), *Nine Characteristics of High-performing Schools: A Research-based Resource for Schools and Districts to Assist with Improving Student Learning*, 2nd Ed., Office of Superintendent of Public Instruction, Olympia, WA, https://files.eric.ed.gov/fulltext/ED499819.pdf. [91]

Shermis, M. and J. Burstein (2013), *Handbook on Automated Essay Evaluation: Current Applications and New Directions*, Routledge. [115]

Shewbridge, C., M. Fuster and R. Rouw (2019), "Constructive accountability, transparency and trust between government and highly autonomous schools in Flanders", *OECD Education Working Papers*, No. 199, OECD Publishing, Paris, https://doi.org/10.1787/c891abbf-en. [23]

Shewbridge, C. et al. (2011), *School Evaluation in the Flemish Community of Belgium 2011*, OECD Reviews of Evaluation and Assessment in Education, OECD Publishing, Paris, https://doi.org/10.1787/9789264116726-en. [13]

Shewbridge, C. and F. Köster (2021), *Promoting Education Decision Makers' Use of Evidence in Flanders*, Educational Research and Innovation, OECD Publishing, Paris, https://doi.org/10.1787/de604fde-en. [5]

Shiel, G., T. Kellaghan and G. Moran (2010), *Standardised Testing in Lower Secondary Education*, [27]

Research Report No. 12, National Council for Curriculum and Assessment, Dublin, https://ncca.ie/media/2454/standardised_testing_in_lowersecondary_education.pdf.

Somers, G. (2021), *Feasibility Study into the Introduction of Central Assessments in Flanders: Legal-technical Aspects (Lot 3)*, Flanders Department of Education and Training, https://data-onderwijs.vlaanderen.be/onderwijsonderzoek/project/1663. [67]

Steunpunt (2022), *Tests for Educational Development: Guidelines for the Central Tests in Flanders*, https://steunpunttoetsen.be/visietekst. [133]

Threlfall, D. and C. Althaus (2021), "A quixotic quest? Making theory speak to practice", in *Learning Policy, Doing Policy: Interactions between Public Policy Theory, Practice and Teaching*, ANU Press, https://doi.org/10.22459/LPDP.2021.02. [137]

Tielemans, K. et al. (2017), *Het Vlaams lager onderwijs in PIRLS 2016. Begrijpend lezen in internationaal perspectief en in vergelijking met 2006 (Flemish Primary Education in PIRLS 2016)*, Centre for Educational Effectiveness and Evaluation, KU Leuven, Leuven, https://data-onderwijs.vlaanderen.be/documenten/bestand.ashx?id=10983. [3]

Timmis, S. et al. (2015), "Rethinking assessment in a digital age: Opportunities, challenges and risks", *British Educational Research Journal*, Vol. 42/3, pp. 454-476, https://doi.org/10.1002/berj.3215. [100]

Torfing, J., E. Sørensen and A. Røiseland (2016), "Transforming the public sector into an arena for co-creation: Barriers, drivers, benefits, and ways forward", *Administration & Society*, Vol. 51/5, pp. 795-825, https://doi.org/10.1177/0095399716680057. [131]

UK Department for Education (2010), *The Importance of Teaching: The Schools White Paper 2010*, HM Government, London, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/175429/CM-7980.pdf. [88]

Van Twist, M. et al. (2013), "Coping with very weak primary schools: Towards smart interventions in Dutch education policy", *OECD Education Working Papers*, No. 98, OECD Publishing Paris, https://doi.org/10.1787/5k3txnpnhld7-en. [55]

Veldkamp, B. and C. Sluijter (eds.) (2019), *Theoretical and Practical Advances in Computer-based Educational Measurement*, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-030-18480-3. [99]

Verger, A., C. Fontdevila and L. Parcerisa (2019), "Reforming governance through policy instruments: How and to what extent standards, tests and accountability in education spread worldwide", *Discourse: Studies in the Cultural Politics of Education*, Vol. 40/2, pp. 248-270, https://doi.org/10.1080/01596306.2019.1569882. [147]

Viennet, R. and B. Pont (2017), "Education policy implementation: A literature review and proposed framework", *OECD Education Working Papers*, No. 162, OECD Publishing, Paris, https://doi.org/10.1787/19939019. [9]

Wandall, J. (2017), "National tests in Denmark: CAT as a pedagogic tool", *FORUM*, Vol. 59/2, p. 145, https://doi.org/10.15730/forum.2017.59.2.145. [69]

Weijts, B. (2023), *Green Light for Flemish Tests: Test with 21,000 Students*, N-VA, [52]

https://www.benweyts.be/nieuws/groen-licht-voor-vlaamse-toetsen-test-bij-21000-leerlingen.

Welsh Government (2022), *Personalised assessments. Guidance for practitioners on individual learner feedback and progress reports.*, https://hwb.gov.wales/api/storage/9ec62e0c-80d9-48c6-95c7-22f681f87c9a/personalised-assessments-guidance-for-practitioners-learner-feedback-sharing-22.pdf. [125]

Welsh Government (2021), *Personalised Assessments: Guidance for Practitioners on Individual Learner Feedback and Progress Report*, Welsh Government, https://hwb.gov.wales/api/storage/408cbf2d-fef4-421a-83bb-b8e7dc14b5e4/211222-personalised-assessments-guidance-for-practitioners-learner-feedback-sharing-with-parents-carers.pdf. [71]

Welsh Government (2019), "HWB - Personalised assessments", https://hwb.gov.wales/curriculum-for-wales/reading-and-numeracy-assessments/personalised-assessments. [124]

Weyts, B. (2019), *Education Policy Note 2019-2024*, https://publicaties.vlaanderen.be/view-file/32245 (accessed on 10 October 2022). [156]

Wyatt-Smith, C., B. Lingard and E. Heck (2021), *Digital Disruption In Teaching And Testing*, Routledge, New York, NY, https://doi.org/10.4324/9781003045793. [101]

Wyse, D. and H. Torrance (2009), "The development and consequences of national curriculum assessment for primary education in England", *Educational Research*, Vol. 51/2, pp. 213-228, https://doi.org/10.1080/00131880902891479. [76]

Annex A. Agenda of OECD team visit to the Flemish Community of Belgium

Monday 13 June	
11h00-12h00	DoET officials responsible for the curriculum and attainment targets (AHOVOKS)
13h00-14h00	Team Vlaamse toetsen – Department Education and Training (DOV)
14h00-15h00	Inspectorate of Education
15h15-16h15	Parent associations (KOOGO)

Tuesday 14 June	
9h00-10h00	Umbrella organisations/PBDs: Catholic schools in Flanders (KOV)
10h00-11h00	Teacher unions
11h00-12h00	Umbrella organisations/PBDs: Community schools (GO!)
13h00-14h00	Umbrella organisations/PBDs: Urban and municipal schools (OVSG)
15h00-16h00	Meeting with the Cabinet

Wednesday 15 June	
9h00-12h00	Arteveldehogeschool, Gent: Information session on the standardised student assessments and discussion groups with teacher students
14h00-16h00	University centre: Steunpunt Centrale Toetsen in Onderwijs, Gent Presentation and interviews with the researchers

Thursday 16 June	
9h00-10h00	DoET officials responsible for communication
10h00-11h00	DoET officials responsible for professional learning (supporting data literacy)
11h15-12h15	Umbrella organisations/PBDs: Provincial schools (POV)
13h30-14h30	Teacher unions
15h00-16h00	Umbrella organisations/PBDs: Small school providers (OKO)

Friday 17 June	
10h00-12h00	Debrief with the DoET and Inspectorate of Education

Annex B. Team members

Dr Michelle Meadows is a Research Fellow at Green Templeton College and Associate Professor in Educational Assessment and Course Director for the MSc in Educational Assessment at the Department of Education, University of Oxford. Michelle has spent her career working at the intersection between research, policy and practice. Prior to joining the university, she was Deputy Chief Regulator and Executive Director of Strategy, Risk and Research at Ofqual, where she conducted research to inform England's educational assessment policy.

Her research evaluates the validity of qualifications and national assessments, including investigations into the setting and maintenance of standards over time, moderation, marking, malpractice, and assessment design. Prior to working at Ofqual she was Director of the AQA's Centre for Education Research and Policy.

Professor Inge de Wolf is a professor in Education Systems at Maastricht University and the Director of Education Lab NL. She connects academics, educational practitioners and policy makers to improve education through academic research. Her research interests are teachers, school quality and equal opportunities in education. Inge also works at the Netherlands Initiative for Education Research, building a knowledge infrastructure for education in the Netherlands. Additionally, she contributes to projects for the OECD and the World Bank, supporting the improvement of various education systems internationally by dealing with targeting factors such as teachers, school development and the use of data.

Claire Shewbridge has been working as an analyst in the OECD Directorate for Education and Skills for over 20 years. She leads the Centre for Educational Research and Innovation's (CERI) work on new professionalism and the future of teaching, which provides a structured methodology for collective future thinking and development of preferred scenarios for the future of the teaching profession. She has developed policy toolkits for self-assessment by decision makers on their use of evidence and peer-learning seminars on strategic education governance. She has worked on education policy reviews, with in-depth work in 13 countries, looking at assessment and evaluation and school resourcing, among others. She co-ordinated secondary analysis of PISA results in the 2000, 2003 and 2006 surveys.

Inés Sanguino is working with the Implementing Education Policies at the OECD Directorate for Education and Skills. Inés is a project manager and analyst for several tailored policy implementation support projects, including with the Flemish Community of Belgium. She previously worked with organisations such as What Works for Children in Social Care and Unlocked Graduates. Most of her work has been in research, collaborating with various projects at the Junior Researcher Programme, King's College London and the University of Oxford, where she also engaged in tutoring undergraduates. Inés completed a BSc in Psychology, and an MPhil in Evidence-Based Social Intervention and Policy Evaluation as a "La Caixa" scholar.

Marco Kools is a project leader and analyst with the OECD Directorate for Education and Skills. He currently leads the Implementing Policies: Leading Education Change programme, which consists of a complex portfolio of implementation support/technical assistance projects, including in the Flemish Community of Belgium, Ireland, Latvia, Moldova, New South Wales (New Zealand) and Spain. He has specialised in various areas of education policy, including the effective policy/programme design and implementation, assessment and evaluation, and (schools as) learning organisations.

In September 2021, Marco returned to the OECD after a two-year secondment with UNICEF Laos where he served as Education Manager of the Partnership for Strengthening the Education System of Lao PDR Project. Before that his work at the OECD was with individual countries such as Latvia, the Netherlands, Sweden and Wales, in support of their school development reform efforts.

Between 2005 and 2012, Marco worked with UNICEF in the Solomon Islands, Laos and at the UNICEF Innocenti Research Centre in Italy. Before that he worked in the field of education in the Netherlands, where he started his career in 1999 as a secondary school teacher. Marco has written and co-ordinated

several publications and academic articles. He holds several degrees, including a PhD in Public Administration, an MBA and a BSc in Educational Sciences.

This Education Policy Perspective has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.