

Pour citer cet article : Dupont, V., Fagnant, A., & Demonty, I. (2026, prépublication). TIMSS 2023 : ce que l'enquête dit (et ne dit pas) des pratiques d'enseignement des mathématiques en 4^e année primaire. *La Revue LEeE*, 11. <https://revue.leeeonline/index.php/info/article/view/277>

TIMSS 2023 : CE QUE L'ENQUÊTE DIT (ET NE DIT PAS) DES PRATIQUES D'ENSEIGNEMENT DES MATHÉMATIQUES EN 4^E ANNÉE PRIMAIRE

Virginie DUPONT, Annick FAGNANT et Isabelle DEMONTY

*Version de la prépublication : août 2026
Évaluation ouverte et collaborative*

Résumé

Au-delà des performances moyennes des élèves et des comparaisons internationales, les enquêtes à large échelle telles que TIMSS offrent des informations riches sur les contextes d'apprentissage dans les différents pays participants. S'appuyant sur le modèle des trois piliers pour un enseignement de qualité (Klieme et al., 2009), cet article analyse les données issues de TIMSS 2023, collectées auprès d'élèves de quatrième année primaire et de leurs enseignant·es, afin de mieux comprendre les pratiques d'enseignement des mathématiques en Fédération Wallonie-Bruxelles. L'analyse repose sur une double approche comparative : d'une part, une mise en perspective avec des systèmes éducatifs plus performants ; d'autre part, une comparaison entre classes au sein de la FW-B selon leur niveau moyen de performance. Les résultats indiquent que les pratiques observées ne diffèrent pas de manière radicale selon les contextes considérés. En revanche, ils permettent de mettre en évidence certaines limites d'interprétation des données ainsi que des pistes d'amélioration pour une exploitation plus fine et nuancée des résultats. L'étude souligne par exemple la nécessité de compléter les données par une prise d'informations au niveau national au travers de questionnaires plus spécifiques sur les pratiques d'enseignement des mathématiques ou de laboratoires cognitifs pour mieux comprendre les difficultés des élèves.

Mots-clés : enquête à large échelle, mathématiques, pratiques d'enseignement

Abstract

Beyond average student performance and international comparisons, large-scale assessments such as TIMSS provide rich information on learning contexts. Drawing on the three basic dimensions of teaching quality framework (Klieme et al., 2009), this article analyzes data from TIMSS 2023, collected from fourth-grade students and their teachers, in order to better understand mathematics teaching practices in the French-speaking Community of Belgium. The analysis is based on a dual comparative approach: on the one hand, it situates the findings in relation to higher-performing education systems; on the other hand, it compares classes within the FW-B according to their average level of achievement. The results indicate that observed practices do not differ radically across the contexts considered. However, they highlight certain limitations in the interpretation of the data, as well as avenues for improvement toward a more refined and nuanced use of these data. The study also emphasizes, for example, the need to complement large-scale assessment data with nationally collected information, through more specific questionnaires on mathematics teaching practices or through cognitive laboratories, in order to better understand students' learning difficulties.

Keywords: large scale assessment, mathematics, teaching practices

1. Introduction

En 2023, la Fédération Wallonie-Bruxelles (FW-B) a participé pour la première fois à l'étude Trends in International Mathematics and Science Study (TIMSS) au niveau de la 4^e année primaire. Organisée sous l'égide de l'Association internationale pour l'évaluation du rendement scolaire (International Association for the Evaluation of Educational Achievement, IEA), cette enquête internationale de grande ampleur vise à évaluer les compétences des élèves en mathématiques et en sciences. Depuis sa création en 1958, l'IEA a joué un rôle majeur dans le développement des enquêtes internationales à large échelle, dont l'importance s'est accrue à partir des années 1990 avec le lancement du premier cycle de TIMSS (Kyriakides et al., 2022).

Comme le soulignent Hastedt et Rocher (2020), l'IEA n'a pas pour mission d'émettre des recommandations normatives, mais bien de fournir aux pays participants un ensemble d'indicateurs décrivant ce que les élèves sont capables de faire dans différents domaines, afin de soutenir l'élaboration de politiques éducatives adaptées aux contextes nationaux. L'utilité de ces indicateurs s'inscrit dans un mode de gouvernance axé sur l'amélioration de l'efficacité des systèmes éducatifs via une gestion par les résultats et la définition d'objectifs quantifiables et mesurables (Bachand & Demers, 2023) : le New Public Management (Maroy, 2021). En fournissant ces indicateurs, les enquêtes internationales à large échelle ont pour objectif de contribuer à l'amélioration de l'enseignement et de l'apprentissage à l'échelle mondiale, en aidant les systèmes éducatifs à mieux comprendre leur propre fonctionnement (Mertes, 2022). En effet, au-delà des classements internationaux qui accompagnent souvent la diffusion des résultats, la perspective comparative offerte par ces enquêtes vise avant tout à encourager les pays à apprendre les uns des autres. Elle permet notamment d'identifier des leviers d'amélioration à partir de l'expérience de systèmes éducatifs confrontés à des défis similaires et évoluant dans des contextes comparables (Mertes, 2022). Par ailleurs, l'organisation cyclique de ces études autorise le suivi de l'évolution des indicateurs dans le temps, tant au niveau national qu'international. Si, au départ, l'étude des pratiques d'enseignement n'était pas l'objectif premier de l'IEA, elle affirme aujourd'hui dans ses missions la volonté de « mieux comprendre les pratiques éducatives, les processus et les politiques, dans le but d'améliorer la qualité de l'enseignement et de l'apprentissage » (<https://www.iea.nl/about/org/mission>).



Cet intérêt accru pour les pratiques d'enseignement se manifeste notamment au travers de la publication d'ouvrages (IEA Research for Educators) et de formats synthétiques à destination des enseignant·es (IEA Teacher Snippets).

Cette perspective est rendue possible grâce au fait que les enquêtes telles que TIMSS offrent bien plus qu'une simple mesure des performances moyennes des élèves (Drent et al., 2013). Elles s'appuient également sur la collecte d'un ensemble riche de données contextuelles, issues de questionnaires adressés aux élèves, aux enseignant·es, aux directions d'école et aux parents, permettant d'éclairer les environnements d'apprentissage (Reynolds et al., 2021). Menées au niveau de classes entières, les évaluations TIMSS offrent en outre la possibilité de relier les caractéristiques de l'enseignement aux apprentissages des élèves (Nilsen & Gustafsson, 2016). Sans prétendre couvrir l'ensemble des dimensions de la qualité de l'enseignement, les données fournies par TIMSS envisagent un spectre particulier : celui des pratiques déclarées par les enseignant·es et des perceptions rapportées par les élèves. Ces matériaux permettent d'éclairer certains aspects de la qualité de l'enseignement, reconnue comme un déterminant majeur des apprentissages (Klieme et al., 2009 ; Seidel & Shavelson, 2007) dans le champ de la recherche sur l'efficacité de l'enseignement (EER ; Scheerens & Bosker, 1997).

Dans ce contexte, la première participation de la FW-B à TIMSS représente une opportunité inédite pour documenter les pratiques d'enseignement des mathématiques au niveau primaire. Le présent article a ainsi pour objectif de questionner la mesure dans laquelle les données issues des questionnaires contextuels de TIMSS, adressés aux élèves de grade 4 et à leurs enseignant·es, permettent d'éclairer ces pratiques. Afin d'y répondre, un ensemble ciblé d'items issus de ces questionnaires est analysé selon une double visée comparative : d'une part, une comparaison avec des pays proches géographiquement et culturellement et affichant de meilleurs résultats, et d'autre part, une comparaison en fonction du profil académique des classes au sein de la FW-B.

2. Revue de la littérature

La qualité de l'enseignement fait référence aux pratiques pédagogiques mises en place par l'enseignant·e au sein de sa classe. Si la conceptualisation de cette qualité de l'enseignement peut varier d'une étude ou d'un pays à l'autre, son caractère multidimensionnel est amplement reconnu (Baumert et al., 2010 ; Kunter et al., 2008 ; Ferguson, 2010 ; Kane & Cantrell, 2012). Les cadres théoriques permettant de conceptualiser cette qualité de l'enseignement peuvent être situés sur un continuum allant des plus génériques aux plus spécifiques (Senden et al., 2021). Les modèles génériques sont ceux qui envisagent la qualité de l'enseignement de manière générale, sans être spécifique à une discipline scolaire en particulier (Creemers et Kyriakides, 2008 ; Klieme et al., 2001 ; Pianta et al., 2008), à l'inverse des cadres conceptuels spécifiques centrés, comme leur nom l'indique, sur une discipline en particulier, par exemple les mathématiques (Hill et al., 2012 ; Kunter et al., 2013). Entre ces deux extrêmes, les modèles hybrides combinent à la fois des éléments issus de cadres conceptuels génériques et spécifiques à une discipline (Charalambous & Praetorius, 2018 ; Schlesinger et al., 2018). Selon Klieme et Nilsen (2022), l'enquête TIMSS mobilise un cadre de référence générique de la qualité de l'enseignement qui est celui des trois piliers fondamentaux développé par Klieme et ses collègues (2009).



2.1. Le modèle des trois piliers pour un enseignement de qualité

Le modèle en trois piliers de la qualité de l'enseignement proposé par Klieme et ses collègues (2009) a émergé à partir de travaux empiriques menés dans le prolongement de l'étude internationale TIMSS-Video de 1995 (Praetorius et al., 2018). Cette étude visait à comparer les pratiques d'enseignement des mathématiques au grade 8 dans différents pays à partir de l'analyse systématique de leçons filmées. Le dispositif international se concentrait sur une description fine des pratiques de classe, telles que la structure des activités, les objectifs pédagogiques ou la nature des tâches proposées, mais ne disposait ni de données sur les élèves ni d'indicateurs de performance, ce qui limitait l'analyse des liens entre enseignement et apprentissage. Pour y remédier, des recherches complémentaires ont été menées en Allemagne et le dispositif a été enrichi de données longitudinales sur les élèves (tests de mathématiques et questionnaires motivationnels administrés en début et en fin d'année scolaire). C'est à partir de ces données que Klieme et ses collègues (2001) ont entrepris une analyse factorielle exploratoire qui a conduit à l'identification de trois dimensions relativement indépendantes mais complémentaires de la qualité de l'enseignement : la gestion de classe, le soutien aux élèves et l'activation cognitive. Chacune de ces dimensions repose par ailleurs sur des théories solides issues de la psychologie et des sciences de l'éducation (Praetorius et al, 2018 ; Lafontaine, 2024).

La gestion de classe est ancrée dans les recherches processus-produits (Lafontaine, 2024) ayant mis en évidence l'efficacité d'un enseignement structuré. Elle repose sur la mise en œuvre de stratégies permettant d'instaurer et de maintenir un environnement ordonné et structuré (Praetorius et al., 2018), permettant d'augmenter au maximum le temps consacré à l'apprentissage (Brunner & Star, 2024). Ce pilier implique à la fois le renforcement des comportements souhaitables par la communication de règles claires, l'établissement de routines stables et la prévention des comportements indésirables grâce à une surveillance constante de la classe et des interventions immédiates si nécessaire. Selon Lafontaine (2024), le caractère structuré de l'enseignement peut également renvoyer à des moments de la leçon durant laquelle l'enseignant·e guide les apprentissages.

Le climat de soutien trouve sa source dans la théorie de l'auto-détermination de Ryan et Deci (2000) selon laquelle les besoins de compétence, d'autonomie et de proximité sociales des élèves doivent être pris en compte pour soutenir leur motivation à apprendre. Cela se manifeste par des interactions chaleureuses et respectueuses entre l'enseignant·e et les élèves, favorisant ainsi de bonnes relations sociales au sein de la classe (Klieme & Nilsen, 2022). Ce pilier englobe à la fois un soutien académique, à travers des explications claires et compréhensibles, et un soutien socio-émotionnel, en accordant une attention individuelle à chaque élève, notamment via des feedbacks constructifs. L'objectif est de créer une atmosphère de confiance et de respect où les élèves se sentent à l'aise pour poser des questions. Pour ce faire, un soutien de qualité de la part de l'enseignant·e suppose une attention réelle aux idées des élèves, la prise en compte de leurs questionnements et la mise à disposition d'une aide supplémentaire lorsque cela s'avère nécessaire. La théorie de Klieme et ses collègues intègrent la clarté de l'enseignement à ce pilier relatif au climat soutenant alors que d'autres auteurs (Kane & Cantrell, 2010 ; Nilsen et al., 2016) font la distinction entre les deux, ajoutant ainsi un quatrième pilier à la qualité de l'enseignement.

Enfin, l'activation cognitive s'appuie sur les théories constructivistes et socio-constructivistes de l'apprentissage qui mettent en évidence le rôle des conflits cognitifs, des liens entre les connaissances antérieures et les nouvelles connaissances ainsi que des interactions sociales dans le processus d'apprentissage (Lafontaine, 2024 ; Praetorius et al, 2018). Ce pilier est



souvent considéré comme plus spécifique à la discipline enseignée, dans la mesure où il requiert la sélection de tâches qui soutiennent la réflexion critique et une résolution de problèmes complexes (Baumert et al., 2010 ; Fauth et al., 2014 ; Klieme et al., 2009). Selon Praetorius et al. (2018), trois dimensions complémentaires structurent ce pilier. Les propos qui suivent les illustrent en mathématiques puisque c'est dans ce domaine que ce concept a été initialement développé. Premièrement, le développement d'une compréhension conceptuelle permet aux élèves d'établir des connexions significatives entre les concepts et les techniques mathématiques. C'est par exemple le cas lorsqu'un·e enseignant·e mobilise une représentation visuelle pour soutenir l'appropriation d'un concept. Deuxièmement, l'exploitation de tâches de haut niveau cognitif, susceptibles de susciter une réflexion approfondie et les conflits socio-cognitifs : en mathématiques par exemple, il peut s'agir de problèmes non routiniers ou de problèmes ouverts présentant plusieurs démarches de résolution. Troisièmement, l'instauration d'un dialogue constructif entre les élèves favorise l'élaboration progressive d'un discours spécifique à la discipline ciblée et une compréhension profonde des concepts. Les enseignant·es peuvent par exemple encourager les discussions en petits groupes, permettant aux élèves d'expliquer leurs démarches ou de confronter leurs idées.

Si ce modèle en trois piliers a initialement été développé en lien avec l'enseignement des mathématiques, il s'est progressivement étendu à d'autres disciplines en raison de son ancrage théorique solide et de son caractère parcimonieux (Praetorius et al., 2018). Il fait par ailleurs l'objet d'un consensus important dans la littérature anglo-saxonne et germanophone (Lafontaine, 2024).

2.2. La cadre de référence TIMSS

Bien que, dès la première enquête TIMSS en 1995, des questionnaires contextuels aient été administrés aux élèves et aux enseignant·es afin de recueillir des informations sur les pratiques de classe, les cadres de référence ne faisaient alors aucune mention explicite de la qualité de l'enseignement (Klieme & Nilsen, 2022). L'accent était principalement mis sur les styles d'enseignement et l'engagement des élèves dans les tâches. Ce n'est qu'au cours des années 2000, alors que la recherche sur la qualité de l'enseignement se développait fortement aux États-Unis et en Allemagne (Baumert et al., 2010 ; Klieme et al., 2009 ; Pianta & Hamre, 2009 cités par Klieme & Nilsen, 2022), que cette thématique a progressivement été intégrée aux cadres de référence mobilisés dans les enquêtes internationales à large échelle de l'IEA.

Ainsi, depuis 2015, l'enquête TIMSS mobilise un cadre de référence générique de la qualité de l'enseignement (Senden et al., 2021) qui, selon Klieme et Nilsen (2022) est celui des trois piliers fondamentaux développé par Klieme et ses collègues (2009). Si ce modèle théorique n'est pas explicitement cité dans le cadre de référence de TIMSS, on y retrouve toutefois les traces des différents piliers (Reynolds et al., 2021).

La gestion de classe est définie dans le cadre de référence de TIMSS 2023 (Reynolds et al., 2021) comme l'ensemble des pratiques qui favorisent l'apprentissage et découragent les comportements perturbateurs. Dans TIMSS, ce sont les élèves qui fournissent les informations relatives à leur perception du climat de classe. Il ne s'agit donc pas ici d'évaluer leur perception des pratiques spécifiques mises en œuvre par leur enseignant·e pour promouvoir les comportements appropriés et limiter les comportements perturbateurs, telles qu'envisagées par Klieme et ses collègues (2009) dans leur définition de la gestion de classe. Le climat soutenant est envisagé à la fois au travers d'éléments relatifs à la clarté de l'enseignement et du soutien à l'apprentissage. En effet, cette dimension est évaluée à partir de la perception qu'ont les élèves de plusieurs aspects : la clarté des explications fournies, la



manière dont l'enseignant·e vérifie leur compréhension, l'utilisation de différentes approches pédagogiques pour répondre à leurs besoins ainsi que la qualité des feedbacks et des réponses apportées à leurs questions. Ces dimensions sont également présentes dans le questionnaire des enseignant·es.

Enfin, l'activation cognitive n'est pas mentionnée explicitement dans le cadre de référence de TIMSS 2023. Ce dernier fait toutefois très brièvement référence à cette dimension en pointant des pratiques telles que demander aux élèves d'expliquer leurs réponses ou encore d'appliquer des procédures mathématiques par eux-mêmes (Reynolds et al, 2021).

3. Objectifs et questions de recherche

Le courant de recherche sur l'efficacité de l'enseignement (Scheerens & Bosker, 1997) a mis en évidence le rôle déterminant de la qualité de l'enseignement pour les apprentissages des élèves (Klieme et al., 2009 ; Seidel & Shavelson, 2007). Les modèles issus de l'EER soulignent en particulier l'importance d'analyser les processus d'enseignement-apprentissage à l'échelle de la classe, en tenant compte à la fois des caractéristiques des enseignant·es, des pratiques mises en œuvre et du contexte d'apprentissage.

Dans cette perspective, les enquêtes internationales à grande échelle telles que TIMSS semblent offrir une opportunité précieuse, en fournissant des données contextualisées permettant d'approcher certaines dimensions des pratiques d'enseignement à partir de questionnaires adressés aux enseignant·es et aux élèves. En FW-B la faiblesse des résultats en mathématiques lors de l'enquête TIMSS 2023 (Demonty et al., 2024 ; Dupont et al., 2024) amène à se questionner sur les pratiques d'enseignement dans cette discipline. Sans prétendre établir des relations causales dont on connaît les limites avec ce type d'enquête (Erikson et al., 2019 ; Hastedt & Rocher, 2020 ; Konstantidinidou & Kyriakides, 2024), l'approche proposée ici vise à mobiliser les données de TIMSS 2023 comme levier pour interroger la qualité de l'enseignement. Ainsi, la question centrale qui guide cette étude est la suivante : dans quelle mesure les données issues des questionnaires contextuels de TIMSS 2023 permettent-elles d'éclairer les pratiques d'enseignement des mathématiques FW-B ? Plus spécifiquement, le cadre mobilisé dans cette enquête présentant des recouvrements avec les trois piliers de Klieme et ses collègues (2009), l'objectif du présent article est d'observer comment se caractérisent les pratiques liées à la gestion de classe, au climat soutenant et à l'activation cognitive en FW-B et d'analyser ce que l'on peut-on en retirer. Pour éclairer cet objectif général, il s'agira de voir dans quelle mesure les pratiques d'enseignement recueillies via l'enquête TIMSS se distinguent dans les pays ou les classes les plus performantes. Plus précisément : Q1 - les pays plus performants déclarent-ils¹, davantage qu'en FW-B, des pratiques considérées dans la littérature comme plus efficaces ? Q2 - Les pratiques déclarées dans les classes les plus performantes en FW-B se rapprochent-elles davantage d'un enseignement de qualité que dans les classes moins performantes ?

¹ Les expressions « les pays déclarent » ou « les classes déclarent » permettent d'exprimer simplement l'idée mais, comme signalé dans le texte, l'article repose sur l'analyse de pratiques déclarées par les enseignant·es et de perceptions des élèves.



4. Méthodologie

4.1. Participant·es

Au printemps 2023, 4 875 élèves de 4^e année primaire de la FW-B ont pris part à l'enquête TIMSS. Ceux-ci sont issus de 165 écoles échantillonnées de manière aléatoire selon des procédures rigoureuses permettant d'assurer la représentativité de l'échantillon et la comparabilité des données de la FW-B avec celles des autres pays auxquels ont été appliquées les mêmes contraintes et précautions (von Davier et al., 2024). Au sein de ces écoles, selon leur taille, une à deux classes ont été tirées au sort aléatoirement. Au total, ce sont ainsi 320 classes de 4^e année primaire qui ont participé à l'enquête. Il est à noter que 365 enseignant·es ont complété un questionnaire contextuel en lien avec l'enseignement des mathématiques. Ce nombre est supérieur aux 320 classes car dans certains cas plusieurs enseignant·es sont chargé·es de l'enseignement des mathématiques dans une même classe. Dans ce cas, chacun·e devait compléter un questionnaire.

Parallèlement, 58 autres pays ont participé à l'enquête permettant ainsi de récolter des données pour 359 098 élèves du grade 4. Dans le présent article, nous analyserons certaines données relatives aux pratiques d'enseignement en FW-B mais aussi dans trois pays performants : l'Angleterre, l'Irlande et les Pays-Bas. Au-delà de leurs bonnes performances en mathématiques, ces pays ont été choisis en raison de la moyenne d'âge des élèves proche de celle de la FW-B et de leur proximité géographique et culturelle. Cette comparaison à des pays culturellement proches est importante car les questions sous forme d'échelles de Likert telles qu'utilisées dans les questionnaires contextuels de l'enquête TIMSS peuvent être sujettes à des biais culturels (Lafontaine, 2017).

4.2. Instrument de mesure

Les données relatives aux pratiques d'enseignement sont issues à la fois des questionnaires contextuels soumis aux élèves et aux enseignant·es. Comme le recommandent Blömeke et ses collègues (2016), les items sélectionnés sont spécifiquement en lien avec l'enseignement des mathématiques tout en retenant, quand cela était possible, des items issus des deux questionnaires afin de refléter à la fois les expériences des élèves et des enseignant·es.

Comme signalé dans le cadre conceptuel de l'enquête TIMSS, la gestion de classe est appréhendée à travers la perception des élèves du climat de classe lors des leçons de mathématiques. Cette dimension, mesurée au moyen de 6 items (tableau 1), constitue une mesure indirecte de la gestion de classe dans la mesure où elle ne porte pas explicitement sur les pratiques des enseignant·es. De plus, nous avons choisi d'intégrer à cette dimension, 3 items issus du questionnaire des enseignant·es (tableau 2). Ces items visent à saisir des pratiques d'enseignement plus « direct », durant lequel l'enseignant·e dirige et structure les concepts et procédures mathématiques.

Tableau 1

Items relatifs au climat de classe issus du questionnaire des élèves

Combien de fois les choses suivantes ont-elles lieu pendant les leçons de mathématiques ? (À chaque leçon ou presque, À environ une leçon sur deux, À quelques leçons, Jamais)
1. Les élèves n'écoutent pas ce que l'instituteur / institutrice dit
2. Il y a trop de bruit, ça empêche les élèves de travailler convenablement



3. L'instituteur / institutrice doit attendre un long moment avant que les élèves se clament
4. Les élèves coupent la parole à l'instituteur / institutrice
5. Les élèves ne respectent pas les règles de la classe
6. Le comportement des autres élèves m'empêche de me concentrer

Tableau 2

Items relatifs à l'enseignement direct issus du questionnaire des enseignant·es

À quelle fréquence demandez-vous aux élèves de faire ce qui suit lorsque vous enseignez les mathématiques à cette classe ? (À chaque leçon ou presque, À une leçon sur deux environ, Parfois, Jamais)

7. M'écouter leur expliquer de nouveaux contenus mathématiques
8. M'écouter leur expliquer comment résoudre des problèmes
9. Travailler tous ensemble en classe sur des problèmes sous ma supervision directe

Concernant le climat soutenant, les items repris dans les questionnaires TIMSS 2023 couvrent aussi bien la notion de soutien académique que celle de clarté. Sept items, issus du questionnaire adressé aux élèves sont en lien avec cette dimension et font explicitement référence à l'enseignement des mathématiques (tableau 3). Dans le questionnaire soumis aux enseignant·es, aucun item ne permet de mesurer les pratiques en lien avec le climat soutenant tout en faisant explicitement référence aux leçons de mathématiques. En effet, les items relatifs à ce pilier (« faire le lien entre la leçon et la vie quotidienne des élèves », « exposer aux élèves les objectifs de la leçon ») sont posés de manière générale (« à quelle fréquence faites-vous ce qui suit lorsque vous enseignez à cette classe ? »). Pour leur caractère trop générique, nous ne traiterons pas ces items dans cet article.

Tableau 3

Items relatifs au climat soutenant issus du questionnaire des élèves

Es-tu d'accord avec les phrases suivantes à propos des leçons de mathématiques ? (Tout à fait d'accord, D'accord, Pas d'accord, Pas du tout d'accord)

1. Mon instituteur/institutrice nous dit clairement ce qu'on doit étudier à chaque leçon
2. Mon instituteur/institutrice est facile à comprendre
3. Mon instituteur/institutrice répond clairement à mes questions
4. Mon instituteur/institutrice explique bien les mathématiques
5. Mon instituteur/institutrice fait beaucoup de choses pour nous aider à apprendre.
6. Mon instituteur/institutrice explique plusieurs fois quelque chose qu'on n'a pas compris.
7. Mon instituteur/institutrice me donne des conseils utiles sur mon travail.

Trois items relatifs à l'activation cognitive seront ici traités. Les deux premiers, issus du questionnaire des élèves, permettent d'illustrer des pratiques où les élèves sont invités à argumenter et à expliquer leurs raisonnements (tableau 4). Un item issu du questionnaire soumis aux enseignant·es (tableau 5) est également illustratif de l'activation cognitive dans le sens où il cible l'établissement de connections entre les connaissances acquises précédemment et les nouvelles connaissances.

Tableau 4

Items relatifs à l'activation cognitive issus du questionnaire des élèves

Es-tu d'accord avec les phrases suivantes à propos des leçons de mathématiques ? (Tout à fait d'accord, D'accord, Pas d'accord, Pas du tout d'accord)

1. Mon instituteur/institutrice me demande de montrer ce que j'ai appris.
2. Mon instituteur/institutrice me demande d'expliquer mes réponses.



Tableau 5

Items relatifs à l'activation cognitive issus du questionnaire des enseignant·es

À quelle fréquence demandez-vous aux élèves de faire ce qui suit lorsque vous enseignez les mathématiques à cette classe ? (À chaque leçon ou presque, À une leçon sur deux environ, Parfois, Jamais)

1. Utiliser par eux-mêmes ce qu'ils ont appris pour résoudre de nouvelles situations

4.3. Analyses

Les analyses de fréquences réalisées sont envisagées ci-dessous selon une double perspective. D'une part, une comparaison internationale est menée avec des pays affichant un meilleur niveau de performance en mathématiques (pour rappel, l'Angleterre, l'Irlande et les Pays-Bas), dans le but d'identifier si des différences significatives existent dans les pratiques déclarées dans ces pays en comparaison avec la FW-B. D'autre part, une analyse différenciée selon le profil académique des classes (performance moyenne) au sein de la FW-B est réalisée, permettant d'examiner dans quelle mesure certaines pratiques sont associées à des contextes d'apprentissage distincts. Ce niveau académique moyen de la classe a été appréhendé en calculant la moyenne obtenue au test TIMSS par les élèves de la classe dans le domaine ciblé. Les classes ont ensuite été catégorisées en 4 quartiles : 25% des classes les plus performantes (Q4), 25% des classes performantes (Q3), 25% des classes moins performantes (Q2) et 25% des classes les moins performantes (Q1). Les pratiques d'enseignement dans les classes académiquement les plus contrastées (Q1 et Q4) ont ensuite été comparées. Pour chacune des pratiques étudiées dans cet article, nous mettrons en évidence celles où des différences significatives apparaissent selon le niveau académique de la classe. Pour statuer sur la significativité des différences de distribution entre les deux groupes (Q1 et Q4), un U de Mann-Witney pour données non-paramétriques a été réalisé (Creswell & Guetterman, 2024). Ce test permet, pour chaque item, de vérifier si les distributions des deux groupes sont significativement différentes ou non.

5. Résultats

5.1. Approche comparative entre pays

5.1.1. Gestion de classe

Les élèves de la FW-B rapportent un climat de classe légèrement plus négatif que celui observé dans les autres pays, tout en étant assez proche de l'Angleterre. Les élèves de FW-B se distinguent particulièrement pour deux items face auxquels ils sont plus de 65% à déclarer la présence de comportements perturbateurs lors d'au moins une leçon sur deux : « Mon institutrice ou mon instituteur doit attendre un long moment avant que les élèves se calment » et « Il y a trop de bruit, ça empêche de travailler convenablement ». Ces pourcentages sont significativement plus faibles dans les 3 pays pris pour comparaison.



Tableau 6

Pourcentage d'élèves qui déclarent que les choses suivantes ont lieu au moins une leçon sur deux lors des leçons de mathématiques

	FW-B	Angleterre	Irlande	Pays-Bas
Les élèves n'écoutent pas ce que l'instituteur / institutrice dit	48	48	45	36***
Il y a trop de bruit, ça empêche les élèves de travailler convenablement	67	59***	44***	42***
L'instituteur / institutrice doit attendre un long moment avant que les élèves se calment	66	60***	49***	55***
Les élèves coupent la parole à l'instituteur / institutrice	52	52	48	43***
Les élèves ne respectent pas les règles de la classe	42	35***	31***	28***
Le comportement des autres élèves m'empêche de me concentrer	53	51	40***	45

Seuil de significativité : *<.05 ; **<.01 ; ***<.001

Les pratiques plus structurantes, où l'enseignant·e explicite des procédures ou de nouveaux contenus sont significativement moins fréquentes en FW-B que dans les pays pris pour comparaison (tableau 7). Par exemple, seul un élève sur deux a un·e enseignant·e qui déclare demander régulièrement (à au moins une leçon sur deux) aux élèves de l'écouter leur expliquer comment résoudre des problèmes alors qu'ils·elles sont environ 9 sur 10 dans ce cas dans les trois pays performants de référence. Des différences importantes, allant de 20 à 30%, se marquent également entre la FW-B et les autres pays de référence pour les deux autres items.

Tableau 7

Pourcentage d'élèves dont l'enseignant·e déclare demander de faire ce qui suit au moins une leçon sur deux lors des leçons de mathématiques

	FW-B	Angleterre	Irlande	Pays-Bas
M'écouter leur expliquer de nouveaux contenus mathématiques	72	95***	96***	99***
M'écouter leur expliquer comment résoudre des problèmes	51	93***	88***	98***
Travailler tous ensemble en classe sur des problèmes sous ma supervision directe	65	91***	84***	84***

5.1.2. Climat soutenant

Les réponses des élèves aux items relatifs au climat soutenant dressent un tableau positif du soutien et de la clarté perçus par les élèves de 4^e année primaire lors des leçons de mathématiques (tableau 8). Pour tous les items, environ 9 élèves sur 10 sont d'accord ou tout à fait d'accord avec le fait que leur enseignant·e est soutenant·e et clair·e dans ses enseignements. Ces tendances sont proches même si significativement plus faibles que celles observées dans les pays pris pour comparaison, à l'exception de la perception de ce que l'enseignant·e met en place pour aider les élèves à apprendre pour laquelle aucune différence significative n'est observée.



Tableau 8

Pourcentage d'élèves d'accord ou tout à fait d'accord avec les propositions suivantes à propos du climat soutenant durant les leçons de mathématiques

	FW-B	Angleterre	Irlande	Pays-Bas
Mon institutrice/instituteur nous dit clairement ce qu'on doit étudier à chaque leçon	91	94***	94***	95***
Mon institutrice/instituteur est facile à comprendre	88	91***	91***	94***
Mon institutrice/instituteur répond clairement à mes questions	86	91***	90***	93***
Mon institutrice/instituteur explique bien les mathématiques	92	93	94**	95***
Mon institutrice/instituteur fait beaucoup de choses pour nous aider à apprendre	94	95	95	95
Mon institutrice/instituteur explique plusieurs fois quelque chose qu'on n'a pas compris	88	91**	93***	93***
Mon institutrice/instituteur me donne des conseils utiles sur mon travail	89	92**	90	88

5.1.3. Activation cognitive

Concernant l'activation cognitive (tableau 9), en FW-B, environ trois quarts des élèves sont d'accord ou tout à fait d'accord avec le fait que leur instituteur ou institutrice leur demande d'expliquer leurs réponses. Cette proportion est légèrement plus faible (71%) lorsqu'il s'agit de montrer ce qu'ils ont appris. Ces pourcentages sont significativement plus faibles que ceux observés dans les pays pris pour comparaison, particulièrement par rapport à l'Angleterre et l'Irlande. Il convient toutefois de rappeler ici le caractère très limité des questions portant sur le concept d'activation cognitive.

Tableau 9

Pourcentage d'élèves d'accord ou tout à fait d'accord avec les propositions suivantes à propos de l'activation cognitive durant les leçons de mathématiques

	FW-B	Angleterre	Irlande	Pays-Bas
Mon institutrice/instituteur me demande d'expliquer mes réponses	75	92***	86***	81***
Mon institutrice/instituteur me demande de montrer ce que j'ai appris	71	83***	79***	74*

L'analyse des données relatives à l'item issu du questionnaire des enseignant-es montre (tableau 10) qu'en FW-B, tout comme dans les pays performants, la majorité des élèves (voire presque la totalité) ont des enseignant-es qui déclarent régulièrement amener leurs élèves à utiliser ce qu'ils ont appris pour résoudre de nouvelles situations. Il est à noter que ce pourcentage, même s'il est d'ampleur importante, est significativement inférieur en FW-B qu'en Angleterre ou aux Pays-Bas.



Tableau 10

Pourcentage d'élèves dont les enseignant·es déclarent demander de faire ce qui suit au moins une leçon sur deux lors des leçons de mathématiques

	FW-B	Angleterre	Irlande	Pays-Bas
Utiliser par eux-mêmes ce qu'ils ont appris pour résoudre de nouvelles situations	84	96***	88	95*

5.2. Approche comparative entre classes au sein de la FW-B

5.2.1. Gestion de classe

Tous les items relevant d'un enseignement plus direct présentent des distributions significativement différentes selon le niveau moyen de performances en mathématiques de la classe (tableau 11).

Tableau 11

Résultats du test U Mann-Witney pour les items relatifs à la gestion de classe

	Statistique	<i>p</i>
M'écouter leur expliquer de nouveaux contenus mathématiques	1056254	0.0002
M'écouter leur expliquer comment résoudre des problèmes	1005187	<.0001
Travailler tous ensemble en classe sur des problèmes sous ma supervision directe	1040294	<.0001

Plus précisément, l'enseignement direct apparaît plus fréquent dans les classes aux performances faibles (tableau 12). Les élèves y sont plus fréquemment invités à écouter l'enseignant·e expliquer de nouveaux contenus ou comment résoudre des problèmes ainsi qu'à travailler ensemble sur des problèmes sous sa supervision directe. Par exemple, 43% des élèves des classes faibles ont un·e enseignant·e qui déclare demander à ses élèves de l'écouter leur expliquer de nouveaux contenus mathématiques à chaque leçon ou presque contre 34% dans ce cas dans les classes au niveau moyen élevé. De manière analogue, environ un tiers des élèves a un·e enseignant·e déclarant demander à ses élèves de l'écouter leur expliquer comment résoudre des problèmes à chaque leçon ou presque alors qu'ils sont un élève sur 5 dans ce cas dans les classes au niveau plus élevé.

Tableau 12

Pourcentages d'élèves dont les enseignant·es déclarent plus ou moins fréquemment des pratiques relevant d'un enseignement direct selon le quartile académique de la classe en mathématiques

M'écouter leur expliquer de nouveaux contenus mathématiques		
	Classes de niveau faible	Classe de niveau élevé
À chaque leçon ou presque	43	34
À une leçon sur deux environ	35	39
Parfois	22	23
Jamais	0	4
M'écouter leur expliquer comment résoudre des problèmes		
À chaque leçon ou presque	34	18
À une leçon sur deux environ	30	29
Parfois	33	45
Jamais	4	8
Travailler tous ensemble en classe sur des problèmes sous ma supervision directe		



À chaque leçon ou presque	33	16
À une leçon sur deux environ	34	51
Parfois	32	29
Jamais	2	5

5.2.2. Climat soutenant

Même si le soutien et la clarté sont perçus comme réguliers par la majorité des élèves, quelle que soit le niveau académique moyen de la classe, des différences significatives selon les quartiles de classe se marquent pour trois items. En effet, l'analyse des distributions (tableau 13) montre qu'il existe une différence significative pour les items relatifs à la clarté des questions, à l'aide apportée pour apprendre et à l'intensité des explications données pour comprendre.

Tableau 13

Résultats du test U Mann-Witney pour les items relatifs au climat soutenant

	Statistique	<i>p</i>
Mon institutrice/instituteur nous dit clairement ce qu'on doit étudier à chaque leçon	1908126	0.1761
Mon institutrice/instituteur est facile à comprendre	1950450	0.1778
Mon institutrice/instituteur répond clairement à mes questions	1960021	0.0059
Mon institutrice/instituteur explique bien les mathématiques	1943657	0.0914
Mon institutrice/instituteur fait beaucoup de choses pour nous aider à apprendre	1938646	0.0319
Mon institutrice/instituteur explique plusieurs fois quelque chose qu'on n'a pas compris	1858282	0.0180
Mon institutrice/instituteur me donne des conseils utiles sur mon travail	1861986	0.3011

Plus précisément, l'analyse des pourcentages (tableau 14) laisse apparaître que les élèves des classes les plus faibles perçoivent l'enseignement comme moins clair et soutenant que les élèves des classes où le niveau moyen est plus élevé. Si les différences de pourcentages sont ténues, il apparaît tout de même que les élèves des classes faibles sont moins nombreux à être tout à fait d'accord avec le fait que leur enseignant·e répond clairement à leurs questions de manière régulière. Ils sont également moins souvent tout à fait d'accord (71% versus 76%) que leur enseignant·e fait beaucoup de choses pour les aider à apprendre.

Tableau 14

Pourcentages d'élèves d'accord ou tout à fait d'accord avec les propositions relevant d'un climat soutenant selon le quartile académique de la classe en mathématiques

Mon institutrice/instituteur répond clairement à mes questions		
	Classes de niveau faible	Classe de niveau élevé
Tout à fait d'accord	53	56
D'accord	28	31
Pas d'accord	15	12
Tout à fait d'accord	4	2
Mon institutrice/instituteur fait beaucoup de choses pour nous aider à apprendre		
Tout à fait d'accord	71	76
D'accord	20	19
Pas d'accord	6	4
Tout à fait d'accord	4	1
Mon institutrice/instituteur explique plusieurs fois quelque chose qu'on n'a pas compris		



Tout à fait d'accord	61	59
D'accord	25	28
Pas d'accord	9	10
Tout à fait d'accord	5	3

5.2.3. Activation cognitive

Concernant l'activation cognitive, les deux items auxquels les élèves ont répondu présentent des distributions significativement différentes selon le niveau de la classe alors que ce n'est pas le cas pour l'item issu du questionnaire des enseignant·es (tableau 15).

Tableau 15

Résultats du test U Mann-Witney pour les items relatifs à la gestion de classe

	Statistique	<i>p</i>
Mon institutrice/instituteur de montrer ce que j'ai appris	1740731	<.0001
Mon institutrice/instituteur me demande d'expliquer mes réponses	1823912	0.0007
Utiliser par eux-mêmes ce qu'ils ont appris pour résoudre de nouvelles situations	1106676	0.1545

Ainsi, du point de vue des élèves, dans les classes les moins performantes, ils sont plus nombreux à être tout à fait d'accord avec le fait que leur enseignant·e leur demande de montrer ce qu'ils·elles ont appris et d'expliquer leurs réponse (tableau 16).

Tableau 16

Pourcentages d'élèves d'accord ou tout à fait d'accord avec les propositions relevant l'activation cognitive selon le quartile académique de la classe en mathématiques

Mon institutrice/instituteur de montrer ce que j'ai appris		
	Classes de niveau faible	Classe de niveau élevé
Tout à fait d'accord	44	33
D'accord	33	33
Pas d'accord	13	21
Tout à fait d'accord	9	13
Mon institutrice/instituteur me demande d'expliquer mes réponses		
Tout à fait d'accord	49	42
D'accord	27	33
Pas d'accord	15	18
Tout à fait d'accord	9	7



6. Discussion et conclusion

Les enquêtes internationales à large échelle occupent aujourd'hui une place importante dans les sciences de l'éducation. Développées dès les années 1990, elles fournissent non seulement des indicateurs de performance, mais également une masse considérable d'informations relatives aux contextes et aux environnements d'apprentissage. Dans cette perspective, l'objectif de cet article était d'exploiter les données issues de l'enquête TIMSS 2023 afin d'éclairer les pratiques d'enseignement des mathématiques en Fédération Wallonie-Bruxelles (FW-B). En s'appuyant sur le modèle des trois piliers de la qualité de l'enseignement proposé par Klieme et al. (2009), l'étude visait plus précisément à proposer un éclairage à la fois descriptif et comparatif sur ces pratiques. Deux approches complémentaires ont ainsi été mobilisées. D'une part, une approche comparative et normative (Erikson et al., 2019) a permis de situer les pratiques observées en FW-B par rapport à celles de systèmes éducatifs plus performants en mathématiques (Angleterre, Irlande, Pays-Bas). D'autre part, afin de dépasser certaines limites des études cross-sectionnelles, notamment l'impossibilité d'établir des relations causales entre pratiques et performances (Konstantinidou & Kyriakides, 2024), une analyse a été conduite en examinant la manière dont ces pratiques varient selon la composition académique des classes, dans la lignée des travaux récents (Nilsen et al., 2025).

Les indicateurs disponibles concernant la gestion de classe suggèrent des environnements d'apprentissage globalement sereins tout en présentant des comportements perturbateurs significativement plus fréquents que dans les pays pris pour comparaison. Cet environnement apparaît par ailleurs moins favorable dans les classes les plus faibles de la FW-B en comparaison avec les classes des niveaux les plus élevés. Des différences encore plus nettes apparaissent en matière de structuration des apprentissages : les pratiques relevant de la structuration des apprentissages par l'enseignant·e, telles qu'écouter l'enseignant·e expliquer de nouveaux contenus ou expliquer comment résoudre des problèmes, sont moins fréquentes en FW-B que dans les pays plus performants pris pour comparaison. En FW-B, cet enseignement plus structuré apparaît par ailleurs comme plus fréquent dans les classes de niveau faible que dans les classes de niveau élevé. Le choix d'intégrer ces items relatifs à un enseignement structuré dans le pilier de la gestion de classe peut être questionnable, notamment en raison des débats que suscite ce type d'enseignement dans le domaine des mathématiques (Sinha, & Kapur, 2021). Par ailleurs, bien que ces éléments ne permettent pas, à eux seuls, de conclure à un déficit de structuration des apprentissages, ils invitent à considérer l'hypothèse d'un effet de désirabilité sociale dans les déclarations des enseignant·es. En effet, dans le contexte des anciens référentiels, les prescriptions curriculaires mettaient fortement l'accent sur la construction par les élèves des notions mathématiques. L'introduction des orientations de la formation mathématique soulignait notamment que « la formation mathématique s'élabore au départ d'objets, de situations vécues et observées dans le réel [...] Le cours de mathématiques ne se limite pas à transmettre des connaissances » (p. 23). Dans un tel cadre normatif, valorisant l'imagination, la réflexion et l'esprit critique, il est plausible que les enseignant·es aient cherché à se conformer à ces attentes institutionnelles en insistant davantage sur l'activation cognitive et la participation des élèves. Pour autant, comme le rappelle le modèle de Klieme et al. (2009), l'efficacité de l'enseignement repose sur une articulation équilibrée entre gestion de classe, activation cognitive et structuration des contenus. Si les deux premiers piliers constituent des leviers essentiels, ils ne sauraient se substituer à des temps d'explicitation des savoirs, des démarches et des stratégies — un enjeu particulièrement crucial pour les élèves les plus en difficulté. Les résultats contrastés en fonction du niveau de la classe amène également à questionner les opportunités d'apprentissage offertes aux élèves des classes les plus faibles et à l'ajustement des pratiques en fonction des élèves présents dans la classe (Schmidt & Maier, 2009). Des attentes moins élevées pourraient en effet conduire à proposer aux élèves



des tâches plus simples et moins engageantes, réduisant ainsi les opportunités d'apprentissage des élèves.

S'agissant du climat soutenant, les résultats indiquent que l'enseignement est généralement perçu comme clair et soutenant. Peu de différences existent avec les pays pris pour comparaison. En revanche, des disparités apparaissent selon le niveau académique des classes : dans les classes les moins performantes, les élèves perçoivent l'enseignement comme moins clair et moins soutenant.

Les résultats relatifs à l'activation cognitive montrent des résultats proches des pays pris pour comparaison tout en étant significativement inférieur. En FW-B, les opportunités d'activation cognitive sont plus fréquentes dans les classes les plus faibles : les élèves perçoivent davantage être invités à expliciter leurs apprentissages ou à mobiliser leurs connaissances dans de nouvelles situations. Ce constat est positif et contrebalance la crainte d'un appauvrissement des opportunités d'apprentissage mentionné ci-dessus. Il est toutefois important de rappeler le débat existant autour du rôle de l'activation cognitive. Si de nombreux travaux ont mis en évidence son effet positif sur les performances (Klieme et al., 2001 ; Baumert et al., 2010 ; Lipowsky et al., 2009 ; Caro et al., 2016), des recherches plus récentes montrent des effets plus nuancés, voire inexistantes ou négatifs (Wang et al., 2023). L'intensité de l'activation cognitive semble jouer un rôle déterminant : elle est bénéfique jusqu'à un certain seuil, au-delà duquel elle peut engendrer une surcharge cognitive et devenir contreproductive (Lyu et al., 2025 ; Liu et al., 2024). En outre, cette dimension apparaît fortement dépendante des contenus disciplinaires, contrairement à la gestion de classe et à la gestion de classe qui constitue une dimension plus générique (Teig et al., 2024 ; Praetorius, 2014, 2018).

Ces différents résultats mettent en lumière la complexité de l'analyse des pratiques d'enseignement et soulèvent davantage de questions qu'elles n'en résolvent. Cette difficulté à dégager des profils de pratiques clairement différenciés invite tout d'abord à interroger les outils mobilisés. Tout d'abord, les items mobilisés dans l'enquête TIMSS ne permettent d'envisager la qualité de l'enseignement que de manière parcellaire. Par exemple, les items relatifs à l'activation cognitive ne sont qu'au nombre de 3 et ne permettent pas de rendre compte des véritables tâches auxquelles les élèves sont confrontés en classe. De plus, certaines dimensions essentielles de l'enseignement, telles que les pratiques d'évaluation, la différenciation pédagogique ou les connaissances professionnelles des enseignant·es, ne sont pas explicitement traitées. Ensuite, le modèle des trois piliers repose sur des dimensions génériques de la qualité de l'enseignement, telles que mobilisées dans TIMSS (Reynolds et al., 2021), et ne permet pas toujours de rendre compte de la spécificité de l'enseignement des mathématiques. Plusieurs travaux récents soulignent la nécessité d'articuler ces dimensions transversales avec des dimensions didactiques propres à la discipline (Blazar et al., 2017 ; Charalambous & Praetorius, 2018 ; Praetorius & Charalambous, 2018 ; Charalambous & Kyriakides, 2017). Enfin, les questionnaires TIMSS renseignent principalement la fréquence des pratiques sans en appréhender la qualité effective, alors même que la fréquence d'une activité ne constitue pas nécessairement un indicateur pertinent de qualité (Blömeke et al., 2016). Certaines limites sont également inhérentes à la manière dont les données ont été traitées dans cet article. Le choix restreint des systèmes éducatifs de comparaison gagnerait à être élargi afin de situer plus finement les résultats. Les données sont essentiellement traitées de manière descriptives et gagneraient certainement à être mises à l'épreuve grâce à des modèles statistiques plus sophistiqués.

Au-delà de ces limites, plusieurs perspectives de recherche se dégagent. Les données TIMSS offrent en effet d'autres ressources encore peu exploitées, telles que les niveaux de compétence (von Davier et al., 2024) ou les items libérés, qui permettent des analyses plus diagnostiques des acquis des élèves, notamment à travers des approches inspirées des



laboratoires cognitifs (Demonty et al., 2024). Les Encyclopédies TIMSS constituent également un outil précieux pour analyser les curricula et les prescriptions nationales. Par ailleurs, le développement de questionnaires nationaux complémentaires permettrait de documenter plus finement certaines dimensions des pratiques d'enseignement, notamment en s'inspirant de travaux sur la résolution de problèmes (Piñeiro et al., 2021). Enfin, l'analyse des pratiques en fonction de la composition des classes apparaît comme une piste particulièrement prometteuse. Encore peu explorée, elle pourrait être approfondie par le recours à des modèles multiniveaux, permettant d'examiner de manière plus rigoureuse les effets conjoints des caractéristiques individuelles et contextuelles (Nilsen et al., 2025).

En conclusion, cet article met en évidence à la fois le potentiel et les limites des enquêtes internationales pour l'étude des pratiques d'enseignement. Si ces données ne permettent pas d'établir des relations causales ni de tirer des conclusions définitives sur l'efficacité des pratiques, elles constituent un levier important pour nourrir la réflexion et formuler des hypothèses. À ce titre, elles contribuent pleinement à l'un des objectifs fondamentaux des enquêtes à large échelle : stimuler la réflexion sur les facteurs susceptibles d'expliquer les différences entre systèmes éducatifs, au-delà des seuls niveaux de performance, et ouvrir la voie à des analyses plus fines, contextualisées et ancrées dans les spécificités disciplinaires.

Références

- Bachand, C.-A., & Demers, S. (2023). Les sciences de l'éducation sur l'autel des données probantes et de l'efficacité ? In S. Chkair & S. Wagnon (Eds.), *Les données probantes et l'éducation* (pp. 47-64). Deboeck.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M. & Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180. <https://doi.org/10.3102/0002831209345157>
- Blazar, D., Braslow, D., Charalambous, C., & Hill, H. (2017). Attending to general and mathematics-specific dimensions of teaching: Exploring factors across two observation instruments. *Educational Assessment*, 22(2), 71–94. <https://doi.org/10.1080/10627197.2017.1309274>
- Blömeke, S., Olsen, R. V., & Suhl, U. (2016). Relation of student achievement to the quality of their teachers and instructional quality. In T. Nilsen & J. E. Gustafsson (Eds.), *Teacher quality, instructional quality and student outcomes* (pp. 21–50). Springer. <https://doi.org/10.1007/978-3-319-41252-8>
- Brunner, E., & Star, J. R. (2024). The quality of mathematics teaching from a mathematics educational perspective: what do we actually know and which questions are still open?. *ZDM–Mathematics Education*, 56(5), 775-787. <https://doi.org/10.1007/s11858-024-01600-z>
- Caro, D. H., Lenkeit, J., & Kyriakides, L. (2016). Teaching strategies and differential effectiveness across learning contexts: Evidence from PISA 2012. *Studies in Educational Evaluation*, 49, 30–41. <https://doi.org/10.1016/j.stueduc.2016.03.005>
- Charalambous, C. Y., & Kyriakides, E. (2017). Working at the nexus of generic and content-specific teaching practices: An exploratory study based on TIMSS secondary analyses. *The Elementary School Journal*, 117(3), 423–454. <https://doi.org/10.1086/690221>
- Charalambous, C. Y., & Praetorius, A.-K. (2018). Studying mathematics instruction through different lenses: Setting the ground for understanding instructional quality more comprehensively. *ZDM: Mathematics education*, 50(3), 355–366. <https://doi.org/10.1007/s11858-018-0914-8>
- Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. Routledge. <https://doi.org/10.4324/9780203939185>



- Creswell, J.W., & Guetterman, D. (2024). *Educational research: Planning, conducting and evaluating quantitative and qualitative research (Seventh Edition)*. Global Edition.
- Demonty, I, Dupont, V., Quittre, V., & Fagnant, A. (2024). *Susciter le raisonnement mathématique des élèves. Pistes d'exploitation d'items de raisonnement issus de l'enquête TIMSS 2023*. Université de Liège. https://www.sciences.uliege.be/cms/c_13556870/fr/timss-focus-sur-raisonner-math?details=true
- Drent, M., Meelissen M. R. M., & van der Kleij, F. M. (2013). The contribution of TIMSS to the link between school and classroom factors and student achievement. *Journal of Curriculum Studies*, 45 (2), 198-224. <https://doi.org/10.1080/00220272.2012.727872>
- Dupont, V., Demonty, I, Quittre, V., & Fagnant, A. (2024). *TIMSS 2023 – Note de synthèse*. Université de Liège. https://www.timss-fwb.uliege.be/upload/docs/application/pdf/2024-12/timss2023-note_de_synthese.pdf
- Erikson, K., Helenius, O., & Ryve, A. (2019). Using TIMSS items to evaluate the effectiveness of different instructional practices. *Instructional Science*, 47, 1-18. <https://doi.org/10.1007/s11251-018-9473-1>
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, 29, 1–9. <https://doi.org/10.1016/j.learninstruc.2013.07.001>
- Hastedt, D., & Rocher, T. (2020). *International large-scale assessments in education: a brief guide*. IEA Compass: Briefs in Education, 10. IEA. <https://www.iea.nl/publications/series-journals/iea-compass-briefs-education-series/september-2020-international-large>
- Hill, C. K., Umland, K., Litke, E., & Kapitula, L. R. (2012c). Teacher quality and quality teaching: Examining the relationship of a teacher assessment to practice. *American Journal of Education*, 118(4), 489–519. <https://doi.org/10.1086/666380>
- Junker, B., Weisberg, Y., Matsumura, L.C., Crosson, A., Wolf, M.K., Levison, A., & Resnick, L. (2006). Overview of the instructional quality assessment (CSE technical report 671). Center for the study of evaluation. National center for research on evaluation, standards, and student testing. <https://cresst.org/wp-content/uploads/R671.pdf>
- Kane, T., & Cantrell, S. (2012). *Gathering feedback for teaching. Combining high-quality observations with student surveys and achievement gains*. MET Project Research Paper, Bill & Melinda Gates Foundation, Seattle, WA. <http://files.eric.ed.gov/fulltext/ED540960.pdf>
- Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: "Aufgabekultur" und Unterrichtsgestaltung im internationalen Vergleich. In E. Klieme & J. Baumert (Eds.), *TIMSS: Impulse für Schule und Unterricht* (pp. 43–47). Bundesministerium für Bildung und Forschung.
- Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Seidel (Eds), *The power of Video studies in investigating teaching and learning in the classrooms* (pp. 137-160). Waxmann.
- Klieme, E. & Nilsen, T. (2022). Teaching Quality and Student outcomes in TIMSS and PISA. In T. Nielsen, A. Stancel-Piatak & J.-E. Gustafsson (Eds.), *International Handbook of Comparative Large-Scale Studies in Education. Perspectives, Methods and Findings*. Springer. <https://link.springer.com/referencework/10.1007/978-3-030-88178-8>
- Konstantinidou, E. & Kyriakides, L. (2024, 13-04). *The Impact of Teaching Quality on Student Cognitive Outcomes: A Follow-Up Study Using TIMSS 2019*. [Paper presentation]. Annual meeting of the American Educational Research Association. Philadelphia, Pennsylvania. <https://doi.org/10.3102/2108141>
- Kunter, M., Tsai, Y.-M., Klusmann, U., Brunner, M., Krauss, S., & Baumert, J. (2008). Students' and mathematics teachers' perceptions of teacher enthusiasm and instruction. *Learning and Instruction*, 18(5), 468–482. <https://doi.org/10.1016/j.learninstruc.2008.06.008>



- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (2013). *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers*. Mathematics Teacher Education, volume 8. Springer. https://doi.org/10.1007/978-1-4614-5149-5_2
- Kyriakides, L., Charalambous, Y., & Charalambous, E. (2022). Using ILSAs to Promote Quality and Equity in Education: The Contribution of the Dynamic Model of Educational effectiveness. In T. Nilsen, A. Stancel-Piatak & J. E. Gustafsson (Eds.), *International Handbook of Comparative Large-Scale Studies in Education* (pp. 253-276). Springer. <https://doi.org/10.1007/978-3-030-88178-8>
- Lafontaine, D. (2017). Évaluations à large échelle : prendre la juste mesure des effets de contexte. In P. Detroz, M. Crahay & A. Fagnant, *L'évaluation à la lumière des contextes et des disciplines* (pp. 21-51). De Boeck Supérieur. <https://doi.org/10.3917/dbu.detro.2017.01.0021>
- Lafontaine, (2024). Vers des modèles éclectiques de la qualité de l'enseignement : comment dépasser les obstacles liés à la coexistence de différents paradigmes de recherche en éducation. In V. Harry, G. Simons & D. Verpoorten (Eds.), *Identifier, modéliser et surmonter les obstacles à l'apprentissage* (pp. 395-403). Presses Universitaires de Liège. <https://books.openedition.org/pulg/46818?lang=fr>
- Lyu, J., Yu, X., Li, S. et al. (2025). The Effect of Students' Perceived Cognitive Activation on Mathematics Performance Depends on the Grade, Performance Level, and Activation Intensity. *International Journal of science and mathematics education*, 23, 2455–2475. <https://doi.org/10.1007/s10763-025-10559-6>
- Liu, X., Hansen, K. Y., Valcke, M., & De Neve, J. (2024). A decade of PISA: students perceived instructional quality and mathematics achievement across European countries. *ZDM–Mathematics Education*, 56(5), 859-891. <https://doi.org/10.1007/s11858-024-01630-7>
- Maroy, C. (2021). L'école Québécoise à l'épreuve de la gestion axée sur les résultats. Sociologie de la mise en œuvre d'une politique néo-libérale. Presses de l'Université de Laval.
- Mertes, N. (2022). Overview of ILSEs and Aspects of Data Reuse. In T. Nilsen, A. Stancel-Piatak & J. E. Gustafsson (Eds.), *International Handbook of Comparative Large-Scale Studies in Education* (pp. 279-301). Springer. <https://doi.org/10.1007/978-3-030-88178-8>
- Nilsen, T., & Gustafsson, J. E. (2016). *Teacher quality, instructional quality and student outcomes*. Springer. <https://doi.org/10.1007/978-3-319-41252-8>
- Nilsen, T., Senden, B., Wangqiong, Y., Jentsch, A., Teig, N., & König, J. (2025). Mathematics teaching quality in classrooms of different compositions in Norway. *ZDM-Mathematics Education*, 57, 1191-1205. <https://doi.org/10.1007/s11858-025-01722-y>
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system: Manual K-3*. Paul H Brookes Publishing.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38, 109–119. <https://doi.org/10.3102/0013189X09332374>
- Piñeiro, J. L., Chapman, O., Castro-Rodríguez, E., & Castro, E. (2021). Prospective Elementary Teachers' Pedagogical Knowledge for Mathematical Problem Solving. *Mathematics*, 9(15), 1811. <https://doi.org/10.3390/math9151811>
- Praetorius, A., Pauli, K., Reusser, C., Rakoczy, K., K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2–12. <https://doi.org/10.1016/j.learninstruc.2013.12.002>
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM*, 50, 407–426. <https://doi.org/10.1007/s11858-018-0918-4>
- Praetorius, A.-K., & Charalambous, C. Y. (2018). Classroom observation frameworks for studying instructional quality: Looking back and looking forward. *ZDM*, 50(3), 535–553. <https://doi.org/10.1007/s11858-018-0946-0>



- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and wellbeing. *American Psychologist*, 55(1), 68-78. <https://psycnet.apa.org/doi/10.1037/0003-066X.55.1.68>
- Reynolds, K. A., Aldrich, C. E. A., Bookbinder, A., Gallo, A., von Davier, M., & Kennedy, A. (Eds.) (2024). *TIMSS 2023 Encyclopedia: Education Policy and Curriculum in Mathematics and Science*. Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.timss.rs5882>
- Rutkowski, D., Rutkowski, L., & von Davier, M. (2014). A brief introduction to modern international large-scale assessment. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 3–9). CRC Press. <https://doi.org/10.1080/15305058.2015.1034867>
- Scheerens, J. & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Pergamon.
- Schlesinger, L., Jentsch, A., Kaiser, G., König, J., & Blomeke, S. (2018). Subject-specific characteristics of instructional quality in mathematics education. *ZDM*, 50(3), 475–490. <https://doi.org/10.1007/s11858-018-0917-5>
- Schmidt, W. H., & Maier, A. (2009). Opportunity to learn. In G. Sykes, B. Schneider, & D.N. Plank (Eds.), *Handbook of Education Policy Research* (pp.541-559). Routledge
- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454–499. <https://doi.org/10.3102/0034654307310317>
- Senden, B., Nilsen, T., & Blömeke, S. (2021). Instructional Quality: A review of Conceptualizations, measurement Approaches and research Findings. In M. Blikstad-Balas, K. Klette & M. Tengberg (Eds.) *Ways of Analyzing Teaching Quality. Potentials and Pitfalls* (pp. 140-172). Scandinavian University Press. <https://doi.org/10.18261/9788215045054-2021-05>
- Sinha, T., & Kapur, M. (2021). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research*, 91(5), 761-798. <https://doi.org/10.3102/00346543211019105>
- Teig, N., Nilsen, T., & Yang Hansen, K. (2024). Theoretical Framework of Teacher Practice. In N. Teig, T. Nilsen, & K. Yang Hansen (Eds.). *Effective and Equitable Teacher Practice in Mathematics and Science Education: A Nordic Perspective Across Time and Groups of Students* (pp. 21–33). Springer. <https://link.springer.com/book/10.1007/978-3-031-49580-9>
- von Davier, M., Fishbein, B., & Kennedy, A. (Eds.). (2024). *TIMSS 2023 Technical Report (Methods and Procedures)*. Boston College, TIMSS & PIRLS International Study Center. <https://timss2023.org/method>
- Wang, X. S., Perry, L. B., Malpique, A., & Ide, T. (2023). Factors predicting mathematics achievement in PISA: A systematic review. *Large-Scale Assessments in Education*, 11(1), 24. <https://doi.org/10.1186/s40536-023-00174-8>

Virginie Dupont

Virginie Dupont est première assistante dans le département des Sciences de l'éducation à l'Université de Liège, Belgique. Ses recherches et ses enseignements portent notamment sur la conception et l'usage des outils d'évaluation. Depuis plusieurs années, elle est impliquée dans les enquêtes PIRLS et TIMSS menées en FW-B.

Annick Fagnant

Annick Fagnant est professeure en Sciences de l'éducation à l'Université de Liège, Belgique. Ses recherches portent notamment sur l'enseignement/apprentissage des mathématiques dans l'enseignement fondamental, sur



Isabelle Demonty

l'apprentissage autorégulé et sur la problématique de l'évaluation en classe. Elle est également impliquée dans plusieurs évaluations internationales telles que TIMSS et TALIS.

Isabelle Demonty est professeure à l'UCLouvain (Belgique), spécialisée en psychopédagogie des apprentissages mathématiques. Elle a contribué aux enquêtes PISA, TIMSS et PIAAC par le développement d'épreuves et l'analyse des résultats en mathématiques. Ses recherches portent sur la résolution de problèmes, la pensée algébrique et l'évaluation au service des apprentissages en mathématiques.

