
TIMSS CM1 2019 : FAIBLESSE DES RÉSULTATS FRANÇAIS, FAIBLESSE DE LA FAMILIARITÉ DES ÉLÈVES AVEC LES ITEMS

Juliette FANJAT¹

Centre national d'étude des systèmes scolaires (Cnesco), CNAM, France

Éric RODITI²

Université Paris Cité, EDA, F-75006 Paris, France

Résumé. En 2019, les élèves scolarisés en CM1 en France ont obtenu de très faibles scores à l'enquête TIMSS, évaluation internationale des acquis des élèves menée par l'*International Association for Evaluation of Educational Achievement* (IEA). Différentes études, mobilisant différentes approches, ont été menées sur ces résultats sans conduire à des explications suffisantes. Nous proposons ici, par une analyse didactique *a priori* de questions évaluatives, de comparer l'évaluation de l'IEA et ce qui est proposé par les manuels scolaires les plus diffusés sur le territoire national. Notre étude, portant sur l'intégralité des exercices proposés dans les rubriques « évaluation » de ces manuels, conduit à mesurer la familiarité des élèves avec les items du questionnaire TIMSS. L'introduction de cette notion de familiarité révèle des écarts importants entre les items TIMSS et les questions évaluatives des manuels scolaires ; elle permet également une amélioration significative de la qualité du modèle statistique expliquant les scores des élèves scolarisés en France en se basant uniquement sur les caractéristiques intrinsèques des items.

Mots-clés. TIMSS, manuels scolaires, évaluation internationale, didactique des mathématiques, CM1.

Introduction

Organisée par l'*International Association for the Evaluation of Educational Achievement* (IEA), l'une des enquêtes *Trends in International Mathematics and Science Study* (TIMSS) cherche à mesurer les performances en mathématiques et en sciences des élèves de la quatrième année de scolarité obligatoire (équivalent du CM1 français). En 2015, la France a participé pour la première fois à cette enquête ; elle a renouvelé sa participation en 2019. Les résultats de cette édition (Mullis *et al.*, 2020) ont fait apparaître la grande faiblesse de la France : en mathématiques, elle se classait dernière de l'Union européenne et avant-dernière des pays de l'Organisation de coopération et de développement économiques (OCDE).

L'objectif de cet article est de proposer une nouvelle piste d'explication des faibles performances des élèves scolarisés en France : les questions auxquelles ces élèves sont entraînés sont peut-être, malgré toutes les précautions prises par les évaluateurs, éloignées de celles qui figurent dans le questionnaire TIMSS. Pour étudier cette hypothèse, il a été choisi de se référer aux manuels scolaires et d'étudier la proximité entre les exercices qu'ils proposent dans les rubriques dédiées à l'évaluation et ceux qui figurent dans l'évaluation de l'IEA. Après avoir justifié la pertinence d'une telle approche, nous présentons le cadre théorique et méthodologique — à la fois

¹ juliette.fanjat@gmail.com

² eric.roditi@u-paris.fr

didactique et psychologique — sur lequel s'appuie notre étude. Nous exposons en quoi les résultats produits permettent d'améliorer la compréhension des scores des élèves scolarisés en France, puis nous concluons notre propos en nous interrogeant sur les limites de cette étude et sur les perspectives qu'elle semble offrir³.

1. Nécessité d'une nouvelle approche

Trois approches du problème que constituent les mauvaises performances de la France en mathématiques ont déjà été mobilisées (voir ci-dessous) : au niveau des enseignants, au niveau des élèves et au niveau de l'organisation générale de l'enseignement de cette discipline à l'école primaire. Comme nous allons le montrer, ces approches s'avèrent pertinentes et pourtant insuffisantes. Cela nous a conduits à adopter une approche centrée sur les exercices d'évaluation auxquels sont soumis les élèves, et particulièrement ceux qui figurent dans les manuels scolaires.

1.1. Trois approches pertinentes mais insuffisantes

La première approche consiste à s'intéresser aux enseignants : niveau de recrutement, formations initiale et continue, satisfaction professionnelle, etc. Sans affirmer l'absence de lien entre ces dimensions et les performances des élèves en mathématiques, il apparaît que se focaliser sur les enseignants ne suffit pas à expliquer directement et immédiatement les résultats des élèves d'un territoire. Par exemple, le travail effectué par Kiwan-Zacka et Piedfer-Quênay (2023) à propos de l'enquête TALIS 2018⁴ montre que les enseignants du primaire japonais sont moins nombreux que leurs homologues français à se déclarer « bien préparés » ou « très bien préparés » à la pédagogie propre aux matières qu'ils enseignent (36,5 % vs 45,7 % - OCDE, 2018). Pourtant, les élèves de CM1 scolarisés au Japon obtiennent les secondes meilleures performances à l'enquête TIMSS (édition 2019), alors que leurs pairs scolarisés en France se classent avant-derniers (Colmant & Le Cam, 2020).

La seconde approche consiste à se focaliser directement sur la population d'enquête (à savoir les élèves eux-mêmes), par exemple en étudiant la corrélation entre les performances des élèves scolarisés dans un pays et leurs conditions socio-économiques. Une note d'analyse du Centre national d'étude des systèmes scolaires (Cnesco) montre, à ce sujet, que la faible réussite des élèves scolarisés en France n'est pas imputable à une surreprésentation d'élèves défavorisés dans ce pays, relativement à ceux de l'OCDE :

La moindre réussite des élèves en France est [...] plus prononcée parmi les élèves défavorisés [...]. En revanche, ce constat ne permet pas d'expliquer à lui seul la faible réussite globale des élèves de CM1 en France relativement aux autres pays de l'OCDE (Botton, 2021, p. 6).

Autrement dit, les faibles performances françaises sont davantage le fait de l'ensemble des élèves que celui d'une catégorie sociodémographique spécifique.

Enfin, la troisième approche, curriculaire, consiste à s'intéresser à l'organisation générale de l'enseignement des mathématiques, par exemple au volume horaire prescrit par les instructions officielles et aux programmes scolaires. Pour le problème qui est le nôtre, ni le volume horaire

³ Cette étude a fait l'objet d'un mémoire de master réalisé par Juliette Fanjat sous la direction d'Éric Roditi. Elle a été menée dans le cadre d'un stage effectué au Centre national d'étude des systèmes scolaires (Cnesco) encadré par Jean-François Chesné. Que celui-ci soit ici remercié pour son accompagnement.

⁴ L'enquête *Teaching and Learning International Survey*, menée par l'OCDE, vise à « décrire les pratiques pédagogiques des enseignants, leur environnement pédagogique et leurs conditions de travail » (Kiwan-Zacka & Piedfer-Quênay, 2023, p. 36).

consacré aux mathématiques ni le programme scolaire ne constituent des pistes satisfaisantes. Les performances de la France à l'enquête TIMSS (édition 2019) sont en effet plus faibles que celles de l'Autriche (485 vs 539) ; pourtant, dans l'enseignement primaire, les élèves ont en France davantage d'heures de mathématiques qu'en Autriche (900 heures soit 20,8 % du temps d'enseignement vs 480 heures soit 17,0 % du temps d'enseignement ; EACEA - Eurydice, 2023). L'étude des programmes scolaires n'est quant à elle pas plus concluante, sans doute car une telle étude ne permet pas de rendre compte de la distance entre curriculum prescrit et curriculum réel, c'est-à-dire entre ce que préconisent les instructions officielles et ce qu'il se passe en classe (Perrenoud, 2010). Dans le cas français par exemple, les attendus de fin de CM1 (2019) consacrent une section entière au domaine de contenu « Données »⁵, comme c'est le cas pour les deux autres que sont « Nombres » et « Géométrie et mesure ». Pourtant, les questions relatives aux données sont celles qui sont le moins bien réussies par la France comparativement aux autres pays de l'OCDE : « [les] sous-scores [des élèves scolarisés en France étant] en retrait significatif par rapport au score global de l'UE » (Colmant & Le Cam, 2020, p. 3).

1.2. Une autre approche : la familiarité avec les questions posées dans TIMSS

Ainsi, les trois approches présentées ci-dessus s'avèrent insuffisantes pour comprendre les faibles performances de la France à l'édition 2019 de l'enquête TIMSS. Notre proposition est de nous intéresser plus directement aux tâches mathématiques effectuées par les élèves en classe, notamment quand ils sont en situation d'évaluation ; ce faisant, nous enrichissons la perspective offerte par l'approche curriculaire. Cette proposition est sous-tendue par une hypothèse : la faiblesse des résultats des élèves scolarisés en France tiendrait aussi au fait qu'ils seraient particulièrement démunis devant les items TIMSS, tant ce qui leur est demandé diffère des tâches qu'ils rencontrent habituellement en évaluation.

La difficulté de recueillir, à grande échelle, les tâches d'évaluation auxquelles les élèves sont réellement confrontés nous a conduits à nous appuyer sur les manuels scolaires. Plus précisément, nous nous sommes appuyés sur les exercices qui y figurent dans les pages dédiées à l'évaluation, et cela pour deux raisons. D'une part, même s'il est peu vraisemblable que les enseignants construisent leurs outils d'évaluation en reproduisant à l'identique les exercices d'évaluation qui figurent dans le manuel qu'ils utilisent, nous faisons l'hypothèse que ces exercices sont représentatifs de ceux que les élèves rencontrent. Cette hypothèse s'appuie notamment sur les résultats de l'enquête « PRAESCO Mathématiques CM2 » concernant les pratiques des enseignants français de CM2 en mathématiques. Cette enquête montre que presque tous les enseignants (98 %) déclarent se référer « souvent » ou « très souvent » à un manuel scolaire au moins pour préparer leurs enseignements de mathématiques (comprenant entre autres la préparation d'exercices à donner en évaluation — même si ces exercices ne sont pas nécessairement choisis parmi ceux des rubriques « évaluation » des manuels) ; pour une majorité d'entre eux (53 %), ce manuel est celui qu'ils utilisent en classe et celui-là seulement (Allard *et al.*, 2022). Notre choix repose d'autre part sur une raison méthodologique. Nous envisageons de développer une étude quantitative des tâches d'évaluation proposées aux élèves en classe, en distinguant ces tâches selon différentes caractéristiques que nous définirons plus loin. La fréquence des tâches analogues constitue selon nous un indicateur de la familiarité potentielle des élèves avec ces tâches dans un contexte d'évaluation. Aussi, retenir toutes celles qui sont présentes dans les manuels exposerait à un risque de surreprésentation de certaines tâches, comme peut-être les tâches techniques qui sont nombreuses dans l'enseignement afin que les élèves puissent suffisamment s'exercer, mais qui n'ont pas de raison de l'être autant dans les

⁵ Celui-ci recouvre les tâches de lecture et de construction de tableaux, de diagrammes et de graphiques.

évaluations.

Nous cherchons donc à savoir si les activités qu'il est nécessaire de mettre en place pour réussir les exercices d'évaluation présents dans les manuels scolaires français de mathématiques sont différentes de celles qui permettent de répondre aux items du questionnaire de l'IEA. Le cas échéant, nous chercherons si un lien peut être établi entre les performances des élèves scolarisés en France à l'enquête TIMSS et leur potentielle familiarité avec les activités requises pour en traiter les items. C'est en effet à partir des exercices d'évaluation auxquels les élèves sont confrontés en classe que se crée leur familiarité avec les items d'un questionnaire d'évaluation — celui de l'évaluation de l'IEA en particulier.

2. La familiarité avec l'évaluation TIMSS, un outil d'analyse

Les objectifs précédemment indiqués nous ont conduits à définir des critères pour analyser des exercices d'évaluation, notamment en fonction des activités mathématiques requises pour y répondre. Ces critères ont permis d'établir une méthode quantitative de comparaison des caractéristiques des exercices d'évaluation, suivant qu'ils figurent dans les manuels scolaires ou dans le questionnaire TIMSS. En tenant compte des différents manuels utilisés en France (puisque c'est avec ces manuels que les enseignants travaillent en classe), les résultats de cette comparaison permettent de rendre compte de la familiarité des élèves avec l'évaluation de l'IEA. Comme nous l'expliquerons en détail, nous avons en fait construit deux indicateurs de familiarité avec cette évaluation. Le premier rend compte de la familiarité des élèves avec les *caractéristiques* des questions TIMSS, tandis que le second rend compte de leur familiarité avec les *items* du questionnaire. Nous chercherons à mettre ce second indicateur en relation avec la réussite aux items, et ainsi à tester notre hypothèse selon laquelle le manque de familiarité avec l'évaluation TIMSS pourrait expliquer la faiblesse des scores des élèves scolarisés en France. Détaillons ces outils méthodologiques après avoir précisé le corpus étudié.

2.1. Définition du corpus d'étude

En ce qui concerne l'enquête internationale, nous n'avons pas eu accès à l'ensemble des items, mais seulement à la moitié d'entre eux (96 sur les 200 que contient le questionnaire). Nous n'avons toutefois *a priori* aucune raison de penser qu'ils sont significativement différents de ceux auxquels nous n'avons pas eu accès. Comme pour les tâches figurant dans les manuels, nous envisageons de développer une étude quantitative des items de l'enquête TIMSS, la fréquence d'items analogues ayant une influence sur le score à l'évaluation.

À propos du choix des manuels scolaires, précisons qu'en France, l'édition de manuels scolaires est libre et qu'il y a de nombreuses collections sur le marché ; cela nous a amenés à ne pas les étudier dans leur intégralité. Pour choisir les manuels scolaires analysés, nous nous sommes appuyés sur le volume des ventes des éditeurs⁶ : nous avons retenu, dans l'ordre décroissant de leur diffusion, les collections permettant de couvrir au moins la moitié des manuels vendus en France. Quatre collections ont suffi pour atteindre cet objectif ; ce sont finalement près de trois mille questions évaluatives présentes dans ces quatre manuels qui ont été analysées.

2.2. Caractérisation des questions évaluatives

Les manuels scolaires présentent des exercices dans leurs pages dédiées à l'évaluation. Ceux-ci

⁶ Ces données sont confidentielles ; aussi, nous ne citerons pas les manuels scolaires étudiés.

sont composés d'un nombre variable de questions unitaires (indépendantes ou non), qualifiées d'items dans le vocabulaire dédié à l'évaluation scolaire et qualifiées de questions évaluatives dans cet article. Aussi mettons-nous en correspondance les *items* de l'évaluation de l'IEA avec les *questions évaluatives* des manuels scolaires⁷.

Attardons-nous à présent sur les critères retenus, à partir de la littérature, pour l'analyse des questions évaluatives des manuels scolaires et des items TIMSS. Nous distinguons à la fois des critères qui ont trait aux *connaissances mathématiques évaluées* (domaine de contenu, domaine cognitif, conversion de registre, contextualisation et niveau de mise en fonctionnement des connaissances) et des critères qui ont trait aux *formats d'évaluation* (nombre de modes d'information de l'énoncé, longueur de l'énoncé textuel et format attendu de réponse). Nous obtenons ainsi huit critères que nous justifions et exemplifions⁸ ci-après, et pour lesquels nous précisons ensuite chacune des modalités.

Les connaissances mathématiques évaluées

Nous appuyant sur la classification de l'IEA, nous distinguons tout d'abord quatre *domaines de contenu* (pour une recension détaillée de ce que recouvrent les trois premiers, voir Lindquist *et al.*, 2017) :

- *Nombres* ;
- *Géométrie et mesure* ;
- *Données* ;
- *Hors de la classification TIMSS* : l'IEA s'efforce de concevoir l'enquête au regard des *curricula* des pays participants. Elle établit une sorte de « *curriculum minimal* », dénominateur commun des différentes instructions officielles⁹. Pour certains pays, il existe donc des parties du curriculum non évaluées par l'IEA : on compte notamment au sein de cette modalité les questions portant sur les nombres négatifs, les pourcentages, l'écriture en chiffres romains, ou encore les notions de hasard et de probabilité.

Notre analyse porte également sur le *domaine cognitif*. Reprenant, là aussi, le cadre de référence de l'enquête TIMSS, ce critère permet de saisir « *le processus de réflexion à évaluer* » (Lindquist *et al.*, 2017, p. 14). Trois domaines cognitifs sont distingués (*ibid.*, p. 18) :

- *Connaître* : ce domaine « *couvre les faits, les concepts et les procédures que les élèves ont besoin de connaître* » :

$$30 \div 6 = ?$$

- *Appliquer* : les questions qui relèvent de cette modalité se concentrent « *sur la capacité des élèves à appliquer leurs connaissances et leur compréhension conceptuelle pour résoudre des problèmes* » :

⁷ Ces deux expressions (*item* et *question évaluative*) désignant des objets équivalents, nous les utilisons de manière indifférenciée dans la suite du texte.

⁸ Pour ce faire, nous utilisons des questions issues ou adaptées des manuels scolaires analysés. En raison de la confidentialité des données sur lesquelles nous nous sommes appuyés, nous ne pouvons toutefois pas citer les manuels scolaires.

⁹ Ce curriculum n'est pas *réellement* minimal : pour chaque pays participant, il existe des items du questionnaire TIMSS « hors programme ». Une « *Test - Curriculum Matching Analysis* » permet à tous les coordinateurs nationaux de s'exprimer sur l'adéquation entre items et instructions officielles et ainsi d'affiner l'analyse des résultats. À ce sujet, voir notamment Mullis *et al.* (2020).

Le professeur a équitablement réparti ses 30 élèves en 6 groupes. Combien d'élèves y a-t-il au sein de chaque groupe ?

- *Raisonnement* : cette modalité « englobe des situations peu familières, des contextes complexes et des problèmes à plusieurs niveaux » (MENJS, 2019b, p. 18) :

Un professeur souhaite répartir 30 élèves de telle manière :

- que chaque groupe ait le même nombre d'élèves, et
- que chaque groupe ait un nombre impair d'élèves.

Trouve une façon permettant au professeur de faire ses groupes.

En mobilisant le cadre de référence de l'IEA, l'introduction des critères « domaine de contenu » et « domaine cognitif » permet d'établir une comparaison directe entre les items TIMSS et les questions évaluatives des manuels scolaires.

Ensuite, nous empruntons à Duval la notion de *conversion de registre*. Celle-ci désigne « la transformation [d'une] représentation en une représentation d'un autre registre en conservant la totalité ou une partie seulement du contenu de la représentation initiale » (Duval, 1995, p. 42) : il s'agit par exemple de passer l'écriture décimale d'un nombre à une écriture sous forme de fraction (de 0,2 à $\frac{1}{5}$). La conversion d'un registre à un autre constitue une activité cognitive en

tant que telle ; de plus, « loin d'être la conséquence de l'acquisition d'un concept mathématique, [elle en] est la condition » (Duval, 2018, p. 17). Son apprentissage suppose une confrontation régulière à des tâches mettant en jeu la coordination de registres sémiotiques (Duval, 2006). Nous distinguons ainsi deux modalités pour ce critère : soit il faut articuler différents registres sémiotiques pour résoudre la question posée (modalité « oui »), soit une telle opération n'est pas nécessaire (modalité « non »). Par exemple, dans l'exemple ci-dessous, l'élève peut lire l'heure pour ensuite procéder au calcul d'horaire, ou encore convertir la durée du trajet en mouvements d'aiguille pour ensuite lire l'heure obtenue. Cette question fait donc appel à une conversion de registre.

Quand Camille est partie de chez elle pour aller à l'école, sa montre indiquait l'heure suivante :



Son trajet a duré 20 minutes. À quelle heure est-elle arrivée ?

La question suivante ne nécessite quant à elle aucune conversion de registre : toutes les informations nécessaires à sa résolution sont présentées dans le registre du langage naturel et peuvent être traitées dans ce registre.

Quand Camille est partie de chez elle pour aller à l'école, sa montre indiquait 8 h. Son trajet a duré 20 minutes. À quelle heure est-elle arrivée ?

Nous prenons également en compte la *contextualisation* de la tâche, qui peut constituer une aide (lorsque la logique du contexte correspond à celle de la tâche mathématique) ou au contraire une charge cognitive supplémentaire, comme le montre l'exemple de l'ascenseur développé par Gellert (2009, p. 123) :

Voici la plaquette accrochée à un ascenseur dans une tour à bureaux.

Cet ascenseur a une capacité de 14 personnes.

En arrivant au bureau le matin, 269 personnes veulent prendre l'ascenseur. Combien d'allers-retours l'ascenseur doit-il faire ?

Theis et Gagnon (2010) reprennent cet exemple et expliquent :

Ce problème est très différent selon qu'il se passe dans la réalité [personnes qui choisissent l'escalier par exemple] ou qu'il devient une tâche scolaire [déterminer le quotient de la division euclidienne de 269 par 14 et y ajouter 1]. On ne demande donc pas à l'élève de traiter la tâche en fonction de ce qu'il se passerait réellement dans le contexte annoncé, mais de traiter la situation selon des critères différents (Theis & Gagnon, 2010, p. 60).

Cela nous amène à penser qu'un élève qui rencontrerait peu d'exercices contextualisés pourrait « passer à côté » des attentes de la tâche, alors même qu'il dispose de toutes les aptitudes nécessaires pour répondre à la question posée. Ainsi, nous distinguons deux modalités pour ce critère : soit la question posée est contextualisée (modalité « oui », qu'importe le type de contexte), soit elle ne l'est pas (modalité « non »).

Enfin, nous nous intéressons au *niveau de mise en fonctionnement des connaissances*. Suivant une proposition de Roditi et Salles (2015) initialement formulée pour analyser les résultats des élèves français au Programme international de suivi des acquis des élèves (PISA), on peut compléter la classification de l'IEA en distinguant différents niveaux « de la mise en fonctionnement des connaissances mathématiques nécessaires pour répondre à la question posée dans l'item » (*ibid.*, p. 239). Nous considérons ainsi quatre niveaux de mise en fonctionnement des connaissances :

- *Concept* : les questions qui relèvent de cette modalité mobilisent le caractère « objet » du savoir mathématique en jeu (Douady, 1986). Les élèves n'ont pas à utiliser ce savoir pour résoudre un problème, mais ils ont à témoigner d'une compréhension qualitative des concepts mobilisés. Dans l'exemple ci-dessous, la focale est placée sur la compréhension des unités de mesure pour elles-mêmes, et pas en tant qu'elles peuvent constituer un outil pour la résolution d'un problème :

Quelle unité est la plus appropriée pour exprimer le périmètre d'un terrain de basket ?

A - Mètre

B - Kilomètre

C - Millimètre

- *Direct* : la question peut être résolue par simple application directe. Dans ce cas, la démarche de résolution est courante, souvent unique, et « les programmes scolaires permettent de penser qu'elle est automatisée pour les élèves » (Roditi & Salles, 2015, p. 240). La question ci-dessous illustre cette modalité :

Un terrain de basket a la forme d'un rectangle de 28 m de longueur sur 15 m de largeur. Quel est son périmètre (en m) ?

En effet, la résolution du problème peut être réduite à l'application de la formule (supposée connue) du calcul du périmètre d'un rectangle ;

- *Adaptation* : la résolution de l'exercice nécessite que « l'élève transforme l'énoncé — les données ou la question posée — avant d'appliquer ses connaissances » (*ibid.*, p. 241). La question ci-dessous, variation de la précédente, illustre cette modalité :

Un terrain de basket a la forme d'un rectangle de 2 800 *cm* de longueur sur 15 *m* de largeur.
Quel est son périmètre (en *m*) ?

En effet, l'élève doit d'abord exprimer en mètres la longueur du terrain (donnée en centimètres) pour calculer le périmètre du rectangle. La conversion est appelée par l'énoncé, qui indique que le périmètre doit être exprimé en mètres.

- *Intermédiaire* : l'élève est amené à prendre des initiatives et doit, « *de manière autonome, introduire un ou plusieurs intermédiaires* » (*ibid.*, p. 241). Considérons ainsi l'exercice ci-dessous :

Le terrain de M^{me} Dupré est un rectangle de 35 *m* de long et 20 *m* de large. Le terrain de M. Coste a la même largeur, mais son périmètre a 30 *m* de plus. Quelle est la longueur du terrain de M. Coste ?

Pour répondre à la question posée, l'élève doit faire preuve d'initiative et introduire des étapes intermédiaires non explicitées par l'énoncé (il peut par exemple calculer les périmètres des rectangles et en déduire la longueur du second, ou introduire que l'accroissement du périmètre correspond à deux fois l'accroissement de la longueur du rectangle).

Le critère « niveau de mise en fonctionnement des connaissances » enrichit le critère « domaine cognitif », leurs modalités n'étant pas parfaitement équivalentes (« direct » n'équivaut par exemple pas à « connaître »). Considérons par exemple l'item TIMSS ci-dessous (d'après Foy *et al.*, 2013) :

Bill is arranging squares in the following way.

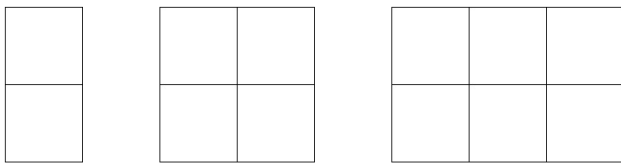


Figure 1 Figure 2 Figure 3

How many squares would Bill need to make Figure 16?

L'IEA classe cet item comme relevant du domaine cognitif « appliquer ». En effet, si l'élève comprend qu'il faut $2n$ carrés pour construire la figure n , alors la résolution de la question n'implique qu'une multiplication par 2. Nous considérons toutefois que cet item relève du niveau de mise en fonctionnement des connaissances « intermédiaire » : à partir des trois premiers éléments de la suite, l'élève a la charge d'en comprendre la règle de construction. Classifier cet item comme relevant du niveau de mise en fonctionnement des connaissances « direct » ou « adaptation » ne rendrait donc pas compte d'une partie non explicite de la tâche, à savoir comprendre qu'au rang n , le nombre de carrés est $2n$.

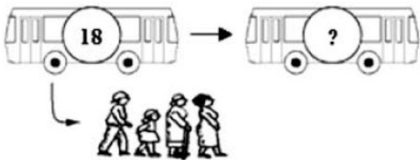
Les formats d'évaluation

Nous retenons trois critères qui ont trait aux formats d'évaluation : le nombre de modes d'information de l'énoncé, la longueur de l'énoncé textuel et le format attendu de réponse.

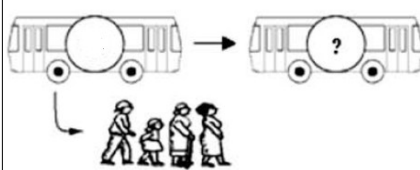
Intéressons-nous tout d'abord au *nombre de modes d'information de l'énoncé*. Un mode d'information peut être textuel, graphique, pictural, etc. Plusieurs modes d'information peuvent être simultanément présents, or la coordination de plusieurs modes d'information non redondants cause un « *effet de dissociation de l'attention* » (Tricot, 1998) ; Sweller *et al.*, (1990, cité par Tricot, 1998) montrent que cette coordination constitue une charge cognitive. Berends et van

Lieshout (2009) montrent quant à eux que la présence conjointe d'un texte et d'une image non-redondante affecte négativement les performances (score et vitesse) des élèves lors de la résolution de problèmes arithmétiques. Pour notre analyse, nous avons distingué deux modalités pour le critère « nombre de modes d'information de l'énoncé » :

- *Un seul mode d'information* (nos exemples sont issus de Berends & van Lieshout, 2009) :

<p><u>Mode d'information numérique</u> :</p> <p>$18 - 4 = ?$</p> <p><u>Mode d'information textuel</u> :</p> <p>Il y avait 18 personnes dans le bus. 4 personnes sont sorties. Combien de personnes sont maintenant dans le bus ?</p> <p><u>Mode d'information pictural</u> :</p> 

- *Plusieurs modes d'information* (*ibid.*) :

<p><u>Modes d'information textuel et pictural</u> :</p> 	<p>Il y avait 18 personnes dans le bus. Des personnes sont sorties. Combien de personnes sont maintenant dans le bus ?</p>
--	--

Dans une perspective analogue, nous nous intéressons également à la *longueur de l'énoncé textuel*. Nous considérons en effet la présence d'un énoncé textuel long comme cognitivement coûteuse, et ce à plus forte raison si les élèves n'y sont pas habitués. À ce titre, la Direction générale de l'enseignement scolaire (Dgesc), dans un guide d'enseignement centré sur la résolution de problèmes au cours moyen, incite les enseignants français à proposer à leurs élèves de tels énoncés :

Il est important de confronter les élèves dès le début du cours moyen à des énoncés de problèmes de plus de deux ou trois lignes sans complexité particulière au-delà de la longueur, afin de leur apprendre à ne pas être déstabilisés par de tels énoncés et à structurer les données pour pouvoir répondre à la question posée (Dgesc, 2022, p. 70).

Nous exprimons la longueur de l'énoncé textuel en nombre de mots, en considérant là aussi trois cas : celui où il n'y a aucun énoncé textuel explicite, celui où le nombre de mots est compris entre 1 et 20 (énoncé textuel court), et celui où le nombre de mots est supérieur ou égal à 21 (énoncé textuel long).

Nous analysons enfin le *format attendu de réponse*. Bodin *et al.* (2016) ont en effet montré que les attendus de réponse de l'enquête TIMSS diffèrent des standards habituels français : là où les élèves français sont habitués à rédiger des phrases complètes, quelques mots ou un dessin peuvent suffire pour obtenir une réponse considérée comme juste selon les consignes de correction de l'IEA. Nous distinguons ainsi quatre modalités pouvant caractériser le format attendu de réponse :

- *Construction* : la réponse à la question nécessite que l'élève trace une figure géométrique, complète un graphique, etc. ;
- *Question à choix multiple*¹⁰ (QCM) : l'élève doit choisir la (ou les) bonne(s) réponse(s) parmi un ensemble de propositions ;
- *Question ouverte à réponse courte* : l'élève peut répondre à la question posée en une phrase ou en quelques mots ;
- *Question ouverte à réponse rédigée* : l'élève doit écrire un texte rédigé pour répondre à la question posée. Notons dès à présent que cette modalité n'est présente dans aucun des 96 items analysés du questionnaire TIMSS.

Bilan : ensemble des critères

L'ensemble de ces considérations conduit à une grille de codage comportant finalement vingt-quatre modalités, qui caractérisent à la fois les connaissances mathématiques évaluées chez les élèves et les formats d'évaluation (*cf.* tableau 1). Deux de ces modalités ne correspondent à aucun item du questionnaire TIMSS ; aussi la grille a-t-elle été réduite à vingt-deux modalités, et nous avons écarté, quand les calculs le nécessitaient, les quelques questions évaluatives des manuels scolaires qui correspondaient à ces deux modalités écartées¹¹.

Critère	Modalité
Connaissances mathématiques évaluées	
Domaine de contenu	Nombres
	Géométrie et mesure
	Données
Domaine cognitif	Connaître
	Appliquer
	Raisonner
Conversion de registre	Oui
	Non
Contextualisation	Oui
	Non
Niveau de mise en fonctionnement des connaissances	Concept
	Direct
	Adaptation
	Intermédiaire

¹⁰ Les items TIMSS auxquels nous avons eu accès ne proposent que des QCM simples (exactement une bonne réponse). Aussi, pour nous adapter au format de traitement des réponses de l'IEA, nous n'avons pas distingué QCM simple et QCM complexe (aucune bonne réponse ou plus d'une bonne réponse). Les QCM complexes à n distracteurs ont donc été traités comme n QCM simples : chaque proposition était considérée comme une question de la forme « vrai / faux ».

¹¹ Le corpus total compte 14 questions évaluatives relevant de la modalité « Question ouverte à réponse rédigée » et 45 questions évaluatives relevant du domaine de contenu « Hors de la classification TIMSS ».

Formats d'évaluation	
Nombre de modes d'information de l'énoncé	Un mode d'information
	Plusieurs modes d'information
Longueur de l'énoncé textuel	Pas d'énoncé textuel
	Entre 1 et 20 mots
	21 mots ou plus
Format attendu de réponse	Question à choix multiple
	Question ouverte à réponse courte
	Construction

Tableau 1 : Critères et modalités considérés pour l'analyse (grille de codage).

Chaque question évaluative se retrouve ainsi caractérisée par huit modalités (une par critère étudié). Dans la suite du texte, nous parlerons donc plus simplement des huit *caractéristiques* de chaque question évaluative (celles du questionnaire TIMSS comme celles extraites des quatre manuels étudiés) afin d'évoquer ces modalités. Deux items sont développés en annexe pour illustrer l'attribution des caractéristiques (l'annexe présente également les indicateurs associés à ces caractéristiques, indicateurs qui font l'objet de la section suivante).

Nous avons finalement codé les 96 items du questionnaire TIMSS auxquels nous avons eu accès et 2 842 questions évaluatives issues des manuels scolaires. Des différences entre les manuels ont été constatées ; nous ne les mentionnons ici pas car notre objectif est bien de tenir compte de l'ensemble des questions évaluatives pour apprécier comment, en moyenne, les élèves scolarisés sont confrontés à ces questions.

Indiquons à présent comment les données recueillies par l'application de cette grille ont été utilisées pour construire les deux indicateurs de familiarité avec l'évaluation de l'IEA.

2.3. Deux indicateurs de familiarité avec l'évaluation TIMSS

Rappelons que le premier indicateur traduit la familiarité des élèves avec les *caractéristiques* des questions de l'évaluation TIMSS et que le second traduit leur familiarité avec les *items* du questionnaire.

Un indicateur de familiarité avec les connaissances évaluées et les formats d'évaluation

Tout d'abord, nous considérons, pour chaque manuel et pour chacune des vingt-deux modalités exposées ci-dessus, la fréquence de cette modalité dans ce manuel :

$$\frac{\text{Nombre de questions évaluatives du manuel relevant de cette modalité}}{\text{Nombre total de questions évaluatives du manuel}}$$

Nous considérons ensuite une fréquence analogue pour le questionnaire TIMSS :

$$\frac{\text{Nombre d'items du questionnaire TIMSS relevant de cette modalité}}{\text{Nombre total d'items du questionnaire TIMSS}}$$

La mise en rapport de ces deux fréquences conduit à définir, pour chaque manuel et pour chacune des vingt-deux modalités exposées ci-dessus, un *indicateur de fréquence relative* dont

l'interprétation est indiquée ci-après (cf. tableau 2) :

Fréquence de la modalité dans le manuel
Fréquence de la modalité dans le questionnaire TIMSS

Valeur de l'indicateur	Interprétation associée
0	Modalité absente des questions évaluatives du manuel.
Inférieure à 1	Modalité proportionnellement <i>moins présente</i> dans les questions évaluatives du manuel que dans le questionnaire TIMSS.
1	Modalité proportionnellement <i>autant présente</i> dans les questions évaluatives du manuel que dans le questionnaire TIMSS.
Supérieure à 1	Modalité proportionnellement <i>plus présente</i> dans les questions évaluatives du manuel que dans le questionnaire TIMSS.

Tableau 2 : Interprétation de l'indicateur de fréquence relative en fonction de sa valeur¹².

Pour chaque manuel analysé, les vingt-deux indicateurs ainsi calculés (un par modalité) indiquent la proximité de ses pages dédiées à l'évaluation avec le questionnaire TIMSS à l'aune de la modalité considérée.

Afin de passer de ces indicateurs concernant les manuels français à des indicateurs concernant les élèves scolarisés en France, nous nous sommes appuyés sur les chiffres de diffusion des manuels : nous considérons en effet que plus un manuel est vendu, plus nombreux sont les élèves qui rencontrent des tâches similaires à celles proposées dans le manuel. Nous avons ainsi, pour chaque modalité, calculé la moyenne des indicateurs de fréquence relative des manuels analysés pondérée par les effectifs des ventes de ces manuels. C'est cette moyenne qui constitue, pour les élèves scolarisés en France, l'*indicateur de leur familiarité avec les connaissances mathématiques évaluées ou le format d'évaluation caractérisé par la modalité considérée*. La familiarité des élèves sera d'autant plus forte que la valeur de l'indicateur de familiarité est élevée.

Nous avons également construit un indicateur de la familiarité des élèves avec les items du questionnaire TIMSS ; nous avons en effet l'objectif d'étudier le lien entre familiarité et scores aux items.

Familiarité des élèves avec les items TIMSS

Les indicateurs de familiarité définis ci-dessus peuvent être utilisés pour rendre compte de la familiarité des élèves scolarisés en France avec un item particulier du questionnaire TIMSS : l'item possédant huit caractéristiques, la familiarité avec l'item peut s'apprécier à l'aune des valeurs des huit indicateurs de familiarité correspondants.

Un indicateur synthétique a été construit pour rendre compte, par une seule valeur numérique, de la familiarité globale des élèves avec l'item. Nous avons calculé la moyenne (géométrique) des indicateurs de familiarité relatifs aux huit caractéristiques de l'item, puis soustrait 1 à la valeur ainsi calculée¹³. Nous obtenons en conséquence, pour chaque item du questionnaire TIMSS, un

¹² Voir le tableau 4 pour un exemple d'interprétation.

¹³ Les indicateurs de fréquence relative étant des rapports de fréquence, ils s'interprètent en fonction de leur valeur

indicateur de familiarité des élèves avec cet item. Le tableau 3 ci-dessous indique l'interprétation de cet indicateur en fonction des valeurs qu'il peut prendre.

Valeur de l'indicateur	Interprétation associée
Inférieure à 0	« Sous-familiarité » des élèves scolarisés en France avec l'item TIMSS.
0	Familiarité neutre des élèves avec l'item TIMSS. La tâche conçue par l'IEA correspond à ce qui est évalué dans les manuels scolaires français : les élèves ne sont pas pénalisés (resp. avantagés) par une sous-familiarité (resp. sur-familiarité) avec l'item.
Supérieure à 0	« Sur-familiarité » des élèves scolarisés en France avec l'item TIMSS.

Tableau 3 : *Interprétation de l'indicateur de familiarité des élèves avec l'item en fonction de sa valeur.*

Pour l'ensemble des items du questionnaire TIMSS, nous avons, afin de tenter d'expliquer le score des élèves scolarisés en France à ces items, effectué une régression multilinéaire¹⁴. Une telle régression permet en effet d'expliquer une variable quantitative — ici les scores des élèves aux items TIMSS — par des variables à la fois quantitatives et qualitatives — ici les caractéristiques intrinsèques aux items d'une part, et la familiarité qu'ont les élèves avec les items d'autre part.

Présentons les résultats produits par cette étude.

3. Familiarité avec l'évaluation TIMSS et scores de la France

Ces résultats s'appuient sur le codage des items du questionnaire TIMSS et des questions évaluatives issues des quatre collections correspondant à plus de la moitié des manuels de mathématiques de CM1 utilisés en France. Nous présentons successivement les indicateurs de familiarité des élèves avec les connaissances mathématiques évaluées et les formats d'évaluation (caractérisés par les vingt-deux modalités retenues dans la grille d'analyse), puis une mise en relation des scores des élèves aux items TIMSS aux indicateurs de leur familiarité avec ces items.

3.1. Une familiarité variable selon les connaissances évaluées et les formats d'évaluation

Les valeurs des indicateurs de familiarité avec les connaissances évaluées et les formats d'évaluation sont présentées dans le tableau 4 ci-dessous. Pour chaque indicateur de familiarité, nous indiquons également le score moyen de la France aux items relevant de la modalité associée, ainsi que la différence avec le score moyen de l'OCDE. Ces éléments mettent en évidence plusieurs éléments de réponse quant à l'existence d'une différence entre les activités induites par les questions évaluatives des manuels scolaires français et celles qu'il faut

par rapport à 1. Nous avons en revanche construit l'indicateur de familiarité de manière à pouvoir, plus simplement, l'interpréter en fonction de son signe plutôt qu'en fonction de sa valeur par rapport à 1.

¹⁴ La régression multilinéaire est une généralisation de la régression linéaire. Ici, elle permet de décrire les variations des scores obtenus par les élèves à chacun des items TIMSS en fonction des différentes variables que sont les caractéristiques des items d'une part et les indicateurs de familiarité des élèves avec les items TIMSS d'autre part.

développer pour réussir les items TIMSS.

Critère	Modalité	Indicateur de familiarité	Score des élèves (France)	Score des élèves (France) – score des élèves (OCDE)
<i>Connaissances mathématiques évaluées</i>				
Domaine de contenu	Nombres <i>N</i> = 1908	1,32**	48,5 %	–10,0 pp
	Géométrie et mesure <i>N</i> = 813	0,83**	46,7 %	–7,1 pp
	Données <i>N</i> = 76	0,18**	47,4 %	–14,2 pp
Domaine cognitif	Connaître <i>N</i> = 2085	1,41**	60,1 %	–6,3 pp
	Appliquer <i>N</i> = 494	0,73**	36,1 %	–12,9 pp
	Raisonner <i>N</i> = 263	0,39**	32,3 %	–13,8 pp
Conversion de registre	Oui <i>N</i> = 924	0,92	44,1 %	–11,1 pp
	Non <i>N</i> = 1918	1,04	49,7 %	–8,8 pp
Contextualisation	Oui <i>N</i> = 579	0,41**	40,3 %	–11,7 pp
	Non <i>N</i> = 2263	1,59**	55,2 %	–7,5 pp
Niveau de mise en fonctionnement des connaissances	Concept <i>N</i> = 126	0,85**	69,5 %	–10,3 pp
	Direct <i>N</i> = 1770	1,07**	53,7 %	–8,1 pp
	Adaptation <i>N</i> = 527	1,98**	40,1 %	–12,0 pp
	Intermédiaire <i>N</i> = 419	0,54**	33,4 %	–11,9 pp

<i>Formats d'évaluation</i>				
Nombre de modes d'information de l'énoncé	Un mode d'information N = 1 980	1,86**	44,8 %	- 11,0 pp
	Plusieurs modes d'information N = 862	0,49**	49,5 %	- 7,7 pp
Longueur de l'énoncé textuel	Pas d'énoncé textuel N = 10	0,06**	59,8 %	- 8,5 pp
	Entre 1 et 20 mots N = 2 514	1,81**	55,4 %	- 7,9 pp
	21 mots ou plus N = 318	0,25**	38,2 %	- 11,8 pp
Format attendu de réponse	Question à choix multiple N = 982	0,61**	55,6 %	- 7,8 pp
	Question ouverte à réponse courte N = 1 726	1,71**	34,9 %	- 11,8 pp
	Construction N = 120	0,81**	48,9 %	- 12,5 pp

Tableau 4 : Valeur des indicateurs de familiarité avec les connaissances mathématiques évaluées et les formats d'évaluation ; score des élèves scolarisés en France et différence de scores avec les élèves scolarisés dans l'OCDE.

Légende : * (resp. **) indique une différence significative au seuil de 5 % (resp. 1 %) entre les questions évaluatives des manuels français et les items du questionnaire TIMSS.

Lecture : l'indicateur de familiarité des élèves correspondant à la modalité « Données » (critère « Domaine de contenu ») est égal à 0,18**. Cela signifie que les manuels français comportent, toutes proportions gardées, 18 questions évaluatives relevant de ce domaine de contenu quand le questionnaire TIMSS en comporte 100, la différence étant significative au seuil de 1 %. Le score des élèves scolarisés en France aux items TIMSS relevant de cette modalité est de 47,4 % ; il est inférieur de 14,2 points de pourcentage à celui des élèves scolarisés dans l'OCDE.

En ce qui concerne le *domaine de contenu*, on constate que les élèves scolarisés en France sont, d'après les manuels scolaires, environ cinq fois moins confrontés en classe à des questions évaluatives relevant du domaine de contenu « Données » que dans le questionnaire TIMSS. Rappelons-le, c'est sur ce domaine que les élèves obtiennent, en France, les scores les plus faibles comparativement aux pays de l'OCDE (- 14,2 points de pourcentage de différence alors qu'elle est de - 10,0 pp et - 7,1 pp pour les deux autres contenus). Toutefois, les élèves sont davantage confrontés en classe aux questions relevant du domaine de contenu « Nombres » (toujours à l'aune de leurs manuels scolaires) que dans le questionnaire TIMSS, mais ce n'est pas le domaine où la différence de score entre la France et les autres pays de l'OCDE est la plus faible (- 10,0 pp alors qu'elle n'est que de - 7,1 pp en « Géométrie et mesure »). En ce qui concerne la *contextualisation*, les résultats font apparaître que la familiarité des élèves scolarisés

en France est plus forte avec les questions sans contextualisation qu'avec les questions contextualisées. C'est également à la fois pour cette modalité que la performance est la plus élevée (62,7 % vs 55,2 %) et que la différence avec les autres pays de l'OCDE est la plus faible (-7,5 pp vs -11,1 pp).

Nous nous sommes aussi intéressés aux critères pour lesquels des modalités différentes témoignent d'un coût cognitif différent de l'activité mathématique de réponse à l'item (domaine cognitif, conversion de registre et niveau de mise en fonctionnement des connaissances) : pour ces trois critères, les résultats synthétisés dans le tableau 4 montrent que les élèves scolarisés en France sont surexposés à la modalité la plus cognitivement accessible et sous-exposés à la modalité la plus cognitivement exigeante (la différence n'étant toutefois pas significative pour le critère conversion de registre).

Détaillons les résultats pour le *domaine cognitif*. L'indicateur de familiarité avec la modalité « Raisonner » — à fort coût cognitif — a pour valeur 0,39. Cette valeur, voisine de 0,4, signifie que, toutes proportions gardées, les élèves scolarisés en France ne rencontrent à l'école que 4 fois cette modalité évaluative (d'après les manuels scolaires) quand elle apparaît 10 fois dans le questionnaire TIMSS. À l'opposé, l'indicateur de familiarité avec la modalité « Connaître » — à plus faible coût cognitif — est égale à 1,4. Cela signifie que les élèves la rencontrent bien davantage dans les évaluations scolaires que dans le questionnaire TIMSS (proportionnellement 14 fois pour 10 fois). Cette différence de familiarité ne semble pas sans lien avec la différence de score entre les items de la modalité « Connaître » (60,1 %) et ceux de la modalité « Raisonner » (32,3 %), ni sans conséquence sur les scores de la France comparativement aux pays de l'OCDE. Ils sont en effet plus faibles encore pour les items relevant de la modalité « Raisonner » (-13,8 pp) que pour ceux de la modalité « Connaître » (-6,3 pp).

Pour ce qui est du *niveau de mise en fonctionnement des connaissances*, les liens entre familiarité et score de la France ne sont pas si clairs. L'indicateur de familiarité avec la modalité « Intermédiaire » est en effet beaucoup plus faible que celui avec la modalité « Adaptation » (0,54 vs 1,98), mais cela ne se traduit pas dans les différences de scores entre la France et les pays de l'OCDE (-11,9 pp vs -12,0 pp).

Nous retenons de ces résultats que les élèves scolarisés en France apparaissent très familiers des questions évaluatives cognitivement peu coûteuses (domaine cognitif « Connaître », un seul mode d'information, énoncé textuel court, question décontextualisée, niveau de mise en fonctionnement des connaissances « Direct ») et peu familiers de celles cognitivement exigeantes (domaine cognitif « Raisonner », au moins deux modes d'information, énoncé textuel long, question contextualisée, niveau de mise en fonctionnement des connaissances « Intermédiaire », etc.). Ce constat pourrait, globalement, expliquer la faiblesse des scores de la France à l'évaluation TIMSS ; il apparaît cependant difficile de l'établir rigoureusement grâce à une méthode reposant sur la familiarité avec les connaissances mathématiques et les formats d'évaluation, et ce à plus forte raison que le travail n'a pas été effectué de manière systématique pour les autres pays de l'OCDE. Cette difficulté nous semble s'expliquer par le fait que les élèves ne sont pas confrontés isolément aux diverses connaissances mathématiques à mettre en œuvre et aux formats variés d'évaluation : ils sont confrontés à des items auxquels correspondent, simultanément, huit modalités qui caractérisent ces connaissances et ces formats d'évaluation. Des effets de renforcement ou de compensation peuvent donc avoir lieu. Toutefois, la familiarité avec les questions évaluatives nous semble constituer un bon candidat pour expliquer les scores des élèves au questionnaire TIMSS.

Pour établir un lien entre familiarité et performance, nous avons par conséquent décidé de poursuivre la recherche en considérant cette fois la familiarité des élèves avec les items TIMSS et le score à ces items.

3.2. La familiarité avec les items contribue à l'amélioration des scores

Nous cherchons donc à savoir dans quelle mesure ces familiarités plus ou moins élevées avec les items sont susceptibles d'expliquer les résultats de la France à l'enquête de l'IEA. L'explication est recherchée statistiquement : la régression multilinéaire montre-t-elle que le score de la France à chacun des items varie en fonction de la familiarité de ses élèves avec ces items ? Avant d'aller plus loin, précisons que la régression multilinéaire permet de traiter simultanément les relations entre les scores aux 96 items et les familiarités à ces items, mais sans perdre les valeurs associées à chacun de ces items.

Les résultats de l'analyse statistique montrent que la réponse à la question que nous posons est clairement positive : les indicateurs de familiarité avec les items TIMSS contribuent, à eux seuls et de manière significative, à l'explication des scores des élèves scolarisés en France. La régression multilinéaire effectuée montre en effet que ces indicateurs rendent compte, pour la France, de plus de 13 % de la variation de performances des élèves aux items du questionnaire TIMSS. Au sein du modèle obtenu, le coefficient associé aux indicateurs de familiarité a une valeur positive ; cela signifie que, pour chacun des items, plus l'indicateur de familiarité des élèves avec l'item est élevé, plus leur score à cet item est élevé également.

Nous avons complété cette étude par une autre régression multilinéaire prenant cette fois en compte, pour expliquer les scores aux items, non seulement les indicateurs de familiarité aux items, mais aussi les caractéristiques des items. Ce travail complémentaire s'avère tout à fait profitable, l'explication des scores de la France étant en effet significativement améliorée¹⁵ : on obtient un modèle qui rend compte de 54 % de la variation des scores des élèves, au lieu des 13 % obtenus en ne tenant compte que des indicateurs de familiarité. L'analyse de ce modèle montre, pour la France, que le score obtenu par les élèves à un item TIMSS dépend finalement, à une constante additive près¹⁶ : du *domaine de contenu* de l'item (nombres, géométrie et mesure, données) ; du *format attendu de réponse* (question ouverte à réponse rédigée, construction ou question à choix multiple), du *niveau de mise en fonctionnement des connaissances* (concept, direct, adaptation ou intermédiaire) et bien sûr de l'*indicateur de familiarité avec l'item*.

L'ensemble de ces résultats répondent à notre question initiale. Nous les discutons ci-après.

4. Discussion

Le travail méthodologique que nous avons mené pour parvenir aux résultats présentés ci-dessus comporte des limites intrinsèques de deux ordres différents, offrant chacune des perspectives quant à notre démarche exploratoire.

4.1. Distinguer familiarité et apprentissage

Notre étude repose sur l'hypothèse que pour résoudre des tâches évaluatives, les élèves ont

¹⁵ Réciproquement, ajouter l'indicateur de familiarité aux modalités caractérisant les items TIMSS quant aux mathématiques évaluées est statistiquement justifié.

¹⁶ En ne conservant que les modalités dont la contribution à l'explication des scores des élèves français est significative ($p < 0,01$).

besoin d'avoir les connaissances requises, mais aussi d'être familiarisés avec ces tâches. On pourrait objecter qu'il y a là une forme de paradoxe : l'élève compétent n'est-il pas justement celui qui, dans une situation nouvelle, est capable de mobiliser ses connaissances pour résoudre la tâche proposée ? Il nous semble cependant que l'hypothèse de notre étude ne doit pas être réduite à l'idée qu'en dehors du « *teaching to the test* », les élèves ne peuvent pas performer, ni même que le moyen à développer pour améliorer les performances de la France soit d'entraîner les élèves sur des questions analogues à celles des évaluations internationales. Nous pensons toutefois que les élèves doivent être familiarisés avec la réalisation de tâches de haut niveau pour être capable d'en réaliser de manière autonome lors d'une évaluation.

Comme il n'y a pas, en France et jusqu'à présent, de « *teaching to the test* », il serait intéressant, pour renforcer notre étude, de mieux contrôler l'hypothèse selon laquelle les enseignants évaluent leurs élèves avec des tâches similaires aux questions évaluatives des manuels scolaires. À cet effet, il serait par exemple intéressant de conduire des enquêtes qualitatives visant à mettre plus précisément en regard les outils d'évaluation des enseignants et les questions évaluatives des manuels. Il serait également intéressant de poursuivre le travail initié dans l'enquête PRAESCO et d'interroger, à grande échelle, le recours des enseignants aux questions des manuels scolaires, notamment pour préparer leurs évaluations.

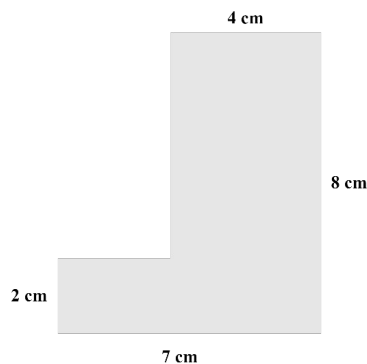
4.2. Hiérarchie des questions évaluatives et effet sur la familiarité

Comme il a déjà été rappelé précédemment, notre étude repose sur la notion de familiarité avec des questions évaluatives, et donc sur l'hypothèse qu'un élève est d'autant mieux préparé à réaliser une tâche qu'il l'a préalablement fréquemment rencontrée. Le fait que les élèves ne réussissent pas à résoudre des problèmes de « haut niveau » (niveau de mise en fonctionnement des connaissances « Intermédiaire », domaine cognitif « Raisonner », etc.) s'ils n'y ont jamais été confrontés est attesté par la littérature (voir par exemple Robert, 1998). Autrement dit, la familiarité avec des questions de « bas niveau » (niveau de mise en fonctionnement des connaissances « Direct », domaine cognitif « Connaître », etc.), si elle favorise la performance à des questions de « bas niveau », n'est pas suffisante pour la réussite des questions de « haut niveau ».

On pourrait en revanche soutenir qu'une question mathématique du domaine cognitif « Raisonner » suppose de « Connaître » un certain nombre de faits mathématiques. De la même manière, résoudre une question avec « Intermédiaire », supposerait de résoudre, *ceteris paribus*, une question où l'emploi des connaissances est « Direct ». Dans cette perspective, on pourrait alors envisager que confronter les élèves à des tâches de haut niveau les familiarise aussi avec celles de plus bas niveau. Illustrons ce propos en considérant l'exemple suivant où l'élève doit calculer un périmètre¹⁷ :

¹⁷ Cet exemple est extrait d'un manuel scolaire de mathématiques hors de notre corpus d'étude — nous ne le citerons pas (voir note de bas de page n° 8).

Calcule le périmètre de cette forme.



Pour répondre à la question posée, l'élève peut calculer la longueur des deux côtés pour lesquels cette longueur n'est pas mentionnée (3 cm pour la longueur horizontale et 6 cm pour la verticale), puis calculer le périmètre de la forme en additionnant les longueurs de tous les côtés. La première sous-tâche met en œuvre implicitement l'additivité de la longueur et constitue un intermédiaire à la charge de l'élève. La tâche est donc de plus haut niveau que si ces deux longueurs avaient été indiquées sur la figure. Une fois les longueurs connues, le périmètre se calcule directement (pas de conversion d'unités par exemple). Ainsi, on pourrait penser que la résolution de cet exercice familiarise *simultanément* les élèves aux questions de « haut niveau » et aux questions de « bas niveau ». Remarquons que l'additivité de la longueur aurait pu être sollicitée en l'absence des indications 4 cm et 2 cm ; la tâche serait alors d'un niveau plus élevé encore. L'indicateur de familiarité, tel que nous l'avons défini, ne rend pas compte de cette simultanéité des familiarisations : il ne rend pas compte de formes de transfert qui pourraient exister depuis les questions de « haut niveau » vers les questions de « bas niveau ».

En l'absence de travaux complémentaires, il convient de rester prudent vis-à-vis de cette hypothèse de transfert. Dans l'exemple précédent, la tâche de plus haut niveau (où seulement les longueurs 7 cm et 8 cm sont indiquées) rend implicite les longueurs des côtés auxquels l'additivité de la longueur s'appliquera ; il n'est donc pas certain que les élèves, surtout les plus faibles, comprennent la résolution du problème. Ils risquent même d'identifier le calcul du périmètre ($2 \times 8 + 2 \times 7$) à l'application de la formule donnant celui du rectangle ($2 \times L + 2 \times \ell$) et de retenir finalement qu'il faut appliquer une formule quand on recherche un périmètre...

Ainsi, si une forme de transfert est possible entre des questions de différents « niveaux », elle apparaît difficilement réductible à un simple processus automatique général. À ce sujet, une étude expérimentale en psychologie cognitive a montré que l'entraînement améliore les performances à un test seulement quand les questions sont d'un niveau similaire :

Initial retrieval practice enhanced final test performance, but only when the initial quiz type (fact or higher order) matched the final test type (fact or higher order, respectively) (Agarwal, 2019, p. 196).

L'indicateur de familiarité, construit et utilisé dans cette étude, ne rend donc pas compte du lien qui unit les tâches de différents « niveaux » ; un lien complexe qu'à notre connaissance, aucun travail scientifique n'a précisément caractérisé.

Conclusion

En 2019, la France a participé pour la seconde fois à l'enquête TIMSS ; les résultats obtenus par les élèves de CM1 ont été largement commentés tant ils furent faibles. Insatisfaits des explications habituellement avancées pour rendre compte de ces performances, nous nous sommes interrogés sur un décalage éventuel entre les tâches auxquelles les élèves scolarisés en France sont habituellement confrontés et celles proposées dans le questionnaire de l'évaluation de l'IEA.

Pour répondre à cette interrogation, nous avons développé et éprouvé une méthodologie pour quantifier la familiarité des élèves avec les items du questionnaire TIMSS. Pour ce faire, nous avons construit des indicateurs reposant sur une analyse didactique *a priori* des questions évaluatives des manuels scolaires, en considérant ces questions comme un « proxy » de celles auxquelles sont confrontés les élèves. Une analyse statistique a mis en lumière des différences entre les items TIMSS et les questions évaluatives offertes par les manuels scolaires diffusés dans plus de la moitié des classes du territoire français. Il ressort de notre étude que les élèves sont très familiers des questions évaluatives cognitivement accessibles et peu familiers des celles qui sont cognitivement exigeantes. Il en ressort également que la prise en compte de la familiarité des élèves avec les items TIMSS aide à comprendre les performances françaises : statistiquement, les scores augmentent avec la familiarité. Il en ressort enfin qu'introduire l'indicateur de familiarité avec les items du questionnaire TIMSS améliore significativement la qualité du modèle statistique d'explication des scores par des variables qui ne portent que sur les items eux-mêmes.

Ajoutons qu'une comparaison internationale, exploratoire seulement, a été effectuée. Elle fait apparaître des différences entre les questions évaluatives des manuels scolaires de différents territoires (Allemagne, Angleterre, Autriche et Irlande). Elle serait à compléter, pour rendre compte de familiarités différenciées entre les élèves européens face aux items TIMSS d'une part, et pour mettre en lien ces éventuelles différences de familiarité avec les différences de scores des différents pays d'autre part.

Il nous semble que la démarche exploratoire que nous proposons, centrée sur les manuels scolaires, gagnerait à être poursuivie et complétée par d'autres approches. Il serait ainsi envisageable de systématiser l'étude des manuels scolaires à l'échelle internationale (notamment en sollicitant des experts connaissant les contextes scolaires et les usages des ressources de chaque territoire), ou encore de la compléter en étudiant les ressources et les outils d'évaluation effectivement proposés aux élèves par les enseignants.

Références bibliographiques

- Agarwal, P. (2019). Retrieval practice & Bloom's taxonomy: Do students need fact knowledge before higher order learning? *Journal of Educational Psychology*, 111(2), 189-209.
<https://doi.apa.org/doiLanding?doi=10.10372Fedu0000282>
- Allard, C., Masselot, P., Peltier-Barbier, M.-L., Roditi, É., Solnon, A. & Tempier, F. (2022). *Résultats de l'enquête sur les pratiques d'enseignement des mathématiques, PRAESCO, en classe de CM2 en 2019*. MEN-DEPP.
<https://hal.science/hal-04146519/file/r-sultats-de-l-enqu-te-sur-les-pratiques-d-enseignement-des-math-matiques-praesco-en-classe-de-cm2-en-2019-118634.pdf>

- Berends, I. E. & van Lieshout E. C. D. M. (2009). The effect of illustrations in arithmetic problem-solving: Effects of increased cognitive load. *Learning and Instruction*, 19(4), 345-353.
<https://doi.org/10.1016/j.learninstruc.2008.06.012>
- Bodin, A., de Hosson, C., Décamp, N., Grapin, N. & Vrignaud, P. (2016). *Analyse comparative des cadres de référence des deux enquêtes (n° 1)*. Paris : Cnesco.
http://www.cnesco.fr/wp-content/uploads/2017/02/161129_RapportPISATIMSS_Vol1.pdf
- Botton, H. (2021). *TIMSS 2019 : des difficultés qui concernent tous les élèves à l'école primaire, plus prononcées parmi les élèves socialement défavorisés*.
<https://www.cnesco.fr/wp-content/uploads/2021/09/Note-analyse-1-comprendre-resultats-mathematiques-France.pdf>
- Colmant, M. & Le Cam, M. (2020). *TIMSS 2019 - Évaluation internationale des élèves de CMI en mathématiques et en sciences : Les résultats de la France toujours en retrait*. Direction de l'évaluation, de la prospective et de la performance (DEPP).
<https://archives-statistiques-depp.education.gouv.fr/Default/digital-viewer/c-47569>
- Douady, R. (1986). Jeux de cadres et dialectique outil-objet. *Recherches en didactique des mathématiques*, 7(2), 5-31.
<https://revue-rdm.com/1986/jeux-de-cadres-et-dialectique/>
- Duval, R. (1995). Registres de représentation sémiotique et fonctionnement cognitif de la pensée. *Annales de didactique et de sciences cognitives*, 5, 37-65.
<https://publimath.univ-irem.fr/numerisation/ST/IST93004/IST93004.pdf>
- Duval, R. (2006). A Cognitive Analysis of Problems of Comprehension in a Learning of Mathematics. *Educational Studies in Mathematics*, 61(1), 103-131.
<https://www.jstor.org/stable/25472062>
- Duval, R. (2018). La conversion des représentations : un des deux processus fondamentaux de la pensée. *Conversion - Du mot au concept* (pp. 9-45). Presses universitaires de Grenoble.
<https://gpc-maths.org/data/documents/duvalconversion.pdf>
- Education, Audiovisual and Culture Executive Agency (EACEA) – Eurydice. (2023). *Recommended annual instruction time in full-time compulsory education in Europe 2022/2023*. Publications Office.
<https://data.europa.eu/doi/10.2797/270455>
- Foy, P., Arora, A. & Stanco, G. M. (éds.). (2013). *TIMSS 2011 released items*.
- Gellert, U. (2009). Zur Explizierung strukturierender Prinzipien mathematischer Unterrichtspraxis. *Journal für Mathematik-Didaktik*, 30(2), 121-146.
<https://doi.org/10.1007/BF03339370>
- Kiwan-Zacka, M. & Piedfer-Quêney, L. (2023). *L'évaluation scolaire : un panorama international des réglementations et des pratiques*. Cnesco-Cnam.

- Lindquist, M., Philpot, R., Mullis, I. V. & Cotter, K. E. (2017). *TIMSS 2019 mathematics framework*. Boston College, TIMSS & PIRLS International Study Center.
<https://timssandpirls.bc.edu/timss2019/frameworks/>
- Mullis, I. V. S., Martin, M. O., Ruddock, G., O'Sullivan, C. Y. & Preuschoff, C. (2020). *TIMSS 2019 international results in mathematics and science*. Boston College, TIMSS & PIRLS International Study Center, Lynch School of Education.
<https://timssandpirls.bc.edu/timss2019/>
- Organisation de coopération et de développement économiques (OCDE) (2018). *Base de données TALIS 2018*.
- Perrenoud, P. (2010). Curriculum réel et travail scolaire. *Travaux de Sciences Sociales*, 227-258.
<https://www.cairn.info/la-fabrication-de-l-excellence-scolaire--9782600014847-page-227.htm>
- Robert, A. (1998). Outils d'analyse des contenus mathématiques à enseigner au lycée et à l'université. *Recherches en didactique des mathématiques*, 18(2), 139-190.
<https://revue-rdm.com/1998/outils-d-analyse-des-contenus/>
- Roditi, É. & Salles, F. (2015). Nouvelles analyses de l'enquête PISA 2012 en mathématiques : un autre regard sur les résultats. *Éducation & formations*, 86-87(11), 235-255.
<https://doi.org/10.48464/ef-86-87-11>
- Sweller, J., Chandler, P., Tierney, P. & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General*, 119, 176-192.
- Theis, L. & Gagnon, N. (2010). Enjeux et limites de la contextualisation en enseignement des mathématiques : points de vue d'une praticienne et d'un chercheur. *Bulletin de l'Association mathématique du Québec*, 50(4), 59-65.
[https://www.amq.math.ca/ancien/bulletins/dec10/Chronique_Primaire\(dec10\).pdf](https://www.amq.math.ca/ancien/bulletins/dec10/Chronique_Primaire(dec10).pdf)
- Tricot, A. (1998). Charge cognitive et apprentissage. Une présentation des travaux de John Sweller. *Revue de Psychologie de l'éducation*, 1, 37-64.
- Dgesco (Direction générale de l'enseignement scolaire) (2022). *La résolution de problèmes mathématiques au cours moyen*.
<https://eduscol.education.fr/document/32206/download?attachment>
- MENJS (2019a). *Attendus de fin de CM1 - Annexe 8 : mathématiques*.
https://cache.media.eduscol.education.fr/file/Attendus_et_reperes_C2-3-4/73/8/08-Maths-CM1-attendus-eduscol_1114738.pdf
- MENJS (2019b). *Grade 4 - Exemples d'items de mathématiques*.
<https://www.education.gouv.fr/media/73330/download>


Annexe



Nous présentons ici l'analyse complète de deux exemples d'item du questionnaire TIMSS, d'après MENJS (2019b).


Exemple 1

Poids des animaux	
Animal	Poids (kg¹⁸)
Guépard	50
Lion	100
Léopard	75

Représente le poids de chaque animal.
Pour cela, fais glisser les symboles nécessaires dans le tableau. Le poids du guépard a déjà été placé pour toi.

Animal	Poids (kg)
Guépard	
Lion	
Léopard	

Code :  = 50 kg

Analysons l'item concernant le poids du lion. 62,8 % des élèves scolarisés en France ont réussi cet item, contre 76,4 % des élèves scolarisés dans l'OCDE. Le tableau ci-dessous récapitule l'analyse que nous proposons de l'item à l'aune des critères que nous considérons.

Critère	Caractéristique	Indicateur de familiarité avec la caractéristique
<i>Connaissances mathématiques évaluées</i>		
Domaine de contenu	Données	0,18
Domaine cognitif	Raisonnement	0,39
Conversion de registre	Oui	0,92

¹⁸ L'énoncé est donné tel qu'il apparaît dans le questionnaire TIMSS, y compris la confusion entre masse et poids qui pourrait faire froncer plus d'un sourcil à des enseignants d'autres disciplines...


	<i>Passage du poids en kg à la représentation sous forme symbolique - 1 poids = 50 kg.</i>	
Contextualisation	Oui	0,41
Niveau de mise en fonctionnement des connaissances	Direct <i>Le poids du lion étant un multiple de 50, l'élève peut directement compléter le tableau en insérant des poids complets ; il n'est ainsi pas nécessaire de comprendre que le demi-poids également présenté représente 25 kg (contrairement à l'item concernant le poids du léopard).</i>	1,07
Formats d'évaluation		
Nombre de modes d'information de l'énoncé	Plusieurs modes d'information <i>3 modes d'information : textuel, pictural, tableau.</i>	0,49
Longueur de l'énoncé textuel	21 mots ou plus <i>26 mots.</i>	0,25
Format attendu de réponse	Construction <i>L'élève doit choisir puis placer le symbole adéquat dans la bonne case.</i>	0,97
Synthèse		
Indicateur de familiarité des élèves scolarisés en France avec l'item		-0,51

Ainsi, les *caractéristiques* de cet item sont les suivantes : données ; raisonner ; oui ; oui ; direct ; plusieurs modes d'information ; 21 mots ou plus ; construction. Les *indicateurs de familiarité des élèves scolarisés en France avec les caractéristiques de cet item* associés sont donc : 0,18 ; 0,39 ; 0,92 ; 0,41 ; 1,07 ; 0,49 ; 0,25 ; 0,97.

Nous calculons la moyenne géométrique de ces huit indicateurs et nous retranchons 1 à ce résultat. Ainsi, après calcul, l'*indicateur de familiarité des élèves scolarisés en France avec cet item* est -0,51. Sa valeur négative indique une « sous-familiarité » des élèves scolarisés en France avec l'item.

Exemple 2

Quelle est la longueur de ce trait, en centimètres ?



A. 7
B. 5,5
C. 3,5
D. 3

40,7 % des élèves scolarisés en France ont réussi cet item, contre 44,6 % des élèves scolarisés dans l'OCDE. Le tableau ci-dessous récapitule l'analyse que nous proposons de l'item à l'aune des critères que nous considérons.

Critère	Caractéristique	Indicateur de familiarité avec la caractéristique
Connaissances mathématiques évaluées		
Domaine de contenu	Géométrie et mesure	0,83
Domaine cognitif	Connaître	1,41
Conversion de registre	Oui <i>Lecture de la longueur du trait sur la règle.</i>	0,92
Contextualisation	Non	1,59
Niveau de mise en fonctionnement des connaissances	Adaptation <i>L'élève doit prendre en compte qu'aucune des extrémités du trait ne correspond à l'origine de la règle.</i>	1,98
Formats d'évaluation		
Nombre de modes d'information de l'énoncé	Plusieurs modes d'information <i>2 modes d'information : textuel et pictural.</i>	0,49
Longueur de l'énoncé textuel	20 mots ou plus <i>9 mots.</i>	1,81
Format attendu de réponse	Question à choix multiple	0,53
Synthèse		
Indicateur de familiarité des élèves scolarisés en France avec l'item		0,06

Ainsi, les *caractéristiques* de cet item sont les suivantes : géométrie et mesure ; connaître ; oui ; oui ; adaptation ; plusieurs modes d'information ; 20 mots ou moins ; question à choix multiple. Les *indicateurs de familiarité des élèves scolarisés en France avec les caractéristiques de cet item* associés sont donc : 0,83 ; 1,41 ; 0,92 ; 1,59 ; 1,98 ; 0,49 ; 1,81 ; 0,53.

Après calcul, l'*indicateur de familiarité des élèves scolarisés en France avec cet item* est 0,06. Sa valeur proche de zéro indique que les élèves scolarisés en France disposent, selon notre modèle, d'une familiarité neutre avec l'item : la tâche conçue par l'IEA correspond à ce qui est évalué dans les manuels scolaires français.