
Élaboration d'un modèle de difficulté de questions évaluant la culture scientifique des élèves

Elaboration of a question difficulty model assessing students' scientific literacy

Mylène Duclos, Florence Le Hebel, Andrée Tiberghien, Pascale Montpied et Valérie Fontanieu



Édition électronique

URL : <https://journals.openedition.org/educationdidactique/9274>
DOI : 10.4000/educationdidactique.9274
ISSN : 2111-4838

Éditeur

Presses universitaires de Rennes

Édition imprimée

Date de publication : 24 novembre 2021
Pagination : 103-131
ISBN : 978-2-7535-8651-2
ISSN : 1956-3485

Référence électronique

Mylène Duclos, Florence Le Hebel, Andrée Tiberghien, Pascale Montpied et Valérie Fontanieu, « Élaboration d'un modèle de difficulté de questions évaluant la culture scientifique des élèves », *Éducation et didactique* [En ligne], 15-3 | 2021, mis en ligne le 03 janvier 2023, consulté le 31 janvier 2026. URL : <http://journals.openedition.org/educationdidactique/9274> ; DOI : <https://doi.org/10.4000/educationdidactique.9274>



Le texte seul est utilisable sous licence CC BY-NC-ND 4.0. Les autres éléments (illustrations, fichiers annexes importés) sont susceptibles d'être soumis à des autorisations d'usage spécifiques.

ÉLABORATION D'UN MODÈLE DE DIFFICULTÉ DE QUESTIONS ÉVALUANT LA CULTURE SCIENTIFIQUE DES ÉLÈVES

Mylène Duclos

École Normale Supérieure de Lyon, ICAR UMR5191, Laboratoire de L'Éducation
LLE UMS 3773, Identifiant ORCID : 0000-0001-9683-754X

Florence Le Hebel

Université de Lyon, École Normale Supérieure de Lyon, ICAR UMR5191,
Laboratoire de L'Éducation LLE UMS 3773,
Identifiant ORCID : 0000-0002-8790-4509

Andrée Tiberghien

CNRS, École Normale Supérieure de Lyon, ICAR UMR5191,
Identifiant ORCID : 0000-0001-8471-1245

Pascale Montpied

CNRS, École Normale Supérieure de Lyon, ICAR UMR5191,
Identifiant ORCID : 0000-0002-7632-3231

Valérie Fontanieu

Institut Français de l'Éducation, École Normale Supérieure de Lyon,
Identifiant ORCID : 0000-0002-4096-597X

Les difficultés des questions posées dans le cadre des évaluations ne sont pas toujours faciles à anticiper par des enseignants ou des concepteurs de tests même si cette expertise est pourtant cruciale. Cet article se concentre sur l'établissement *a priori* des caractéristiques des questions qui expliqueraient les difficultés des élèves, difficultés dont certaines possiblement spécifiques du milieu socio-économico-culturel et du niveau de performance des élèves. Notre étude a donc consisté à élaborer un modèle prédictif des difficultés de questions relevant de la culture scientifique. Pour cela, nous prenons l'exemple des tâches de culture scientifique de l'enquête internationale PISA 2015. Cela nous permet de discuter la construction de notre modèle, les attentes de résultats qu'il peut induire et l'implication de ce type d'études pour l'enseignement et la conception d'évaluations en sciences.

Mots-clés : culture scientifique, modèle de difficulté de question, statut socio-économico-culturel des élèves, performance PISA Science 2015.

Elaboration of a question difficulty model assessing students' scientific literacy

The difficulties of the questions asked in the framework of the assessments are not always easy to anticipate by teachers or test designers even if this expertise is nevertheless crucial. This article focuses on establishing a priori of the characteristics of the questions that would explain the students' difficulties, some of which may be specific to the socioeconomic-cultural environment and the level of performance of the students. Our study therefore consisted in developing a predictive model of the difficulties of questions related to scientific literacy. For this, we choose the example of the scientific literacy tasks of the international PISA 2015 survey. This allows us to discuss the construction of our model, the expectations of outcomes it might induce and the implication of this type of studies for teaching and designing assessments in science.

Keywords: scientific literacy, model of question difficulty, socioeconomic-cultural status of students, PISA Science 2015 performance.

INTRODUCTION

Dans le domaine de l'évaluation, la compréhension de ce qui rend une question difficile pour les élèves est cruciale autant pour les enseignants que pour les concepteurs d'évaluations de tests (El Masri *et al.*, 2017). Différentes études ont montré que les enseignants peuvent majoritairement anticiper le succès ou l'échec, en d'autres termes la performance de leurs élèves sur un contenu défini (Maurice, 1996 ; Südkamp *et al.*, 2012 ; Fakcharoenphol *et al.*, 2015). En revanche, les prédictions des enseignants sont nettement moins bonnes sur la façon dont les élèves résolvent les questions. Nos recherches antérieures vont dans le même sens ; elles ont porté sur la capacité des enseignants de sciences (Physique/Chimie ; SVT) et de technologie en France à prédire les difficultés rencontrées par les élèves moyennement à faiblement performants, lors de la résolution de tâches de sciences, en prenant le cas des items PISA Science 2006 (Le Hebel *et al.*, 2019). Nos résultats ont montré que, même si les enseignants font preuve d'une bonne capacité à prédire les niveaux de difficulté des items pour les élèves moyennement à faiblement performants, ils ne peuvent prédire la façon dont les élèves résolvent les questions et n'identifient pas certains éléments essentiels des difficultés rencontrées par les élèves lors de la résolution de ces items (Le Hebel *et al.*, 2019).

Notre étude se situe dans la perspective d'améliorer l'expertise des enseignants et des concepteurs d'évaluation sur la prédiction des difficultés de questions relevant de la culture scientifique. Nous cherchons à comprendre quelles caractéristiques de questions de sciences influencent les écarts des performances des élèves selon leur statut socio-économique et culturel et leur niveau de performance. Ce niveau de performance correspond au score obtenu par les élèves à une évaluation comportant un ensemble de questions.

La composante du milieu socio-économique et culturel nous semble actuellement essentielle comme le montrent aussi bien les résultats des enquêtes sur des échantillons représentatifs d'élèves comme PISA (Programme International pour le Suivi et les Acquis des élèves) (OCDE, 2016) ou TIMMS (Trends in International Mathematics and Science Study) (IEA, 2016) que des travaux de sociolo-

gie avec des approches qualitatives (Lahire, 1995 ; Bonnery, 2007 ; Bautier et Rayou, 2013). D'une part, les résultats de PISA 2015 montrent des scores pour la France dans la moyenne des pays de l'OCDE et surtout l'une des plus fortes corrélations des performances des élèves avec leur Statut Économique Social et Culturel (SESC), comparé aux autres pays de l'OCDE. D'autre part, des sociologues ont analysé le fonctionnement des élèves de faible SESC et certains (Lahire, 1995 ; Bautier et Rayou, 2013 ; Bonnery, 2007) soulignent la nécessité de s'intéresser plus précisément aux processus qui amènent les élèves à répondre différemment aux tâches lorsqu'ils viennent d'un milieu social différent, afin de permettre aux enseignants d'appréhender de manière plus adéquate les difficultés rencontrées par les élèves.

Pour cette étude nous avons choisi de prendre le cas des items PISA Sciences 2015 évaluant les connaissances et les compétences, chez les élèves de quinze ans, essentielles à leur vie en société (OCDE, 2016) (2015 étant l'année où la culture scientifique est le domaine majeur, c'est-à-dire domaine qui présente le plus de questions). Chaque item PISA comprend une introduction suivie d'une question.

En effet, comme nous l'avons mentionné précédemment, outre les questions relevant de la culture scientifique, cette évaluation comporte des informations sur le Statut Économique Social et Culturel (SESC) des élèves. Elle permet donc d'avoir des résultats sur un échantillon suffisamment important pour comparer les performances des élèves selon leur statut socio-économique et culturel et/ou suivant leur niveau de performance.

Après une présentation de notre cadre théorique sur les modèles prédictifs de difficultés d'une question, et l'élaboration de notre propre modèle prédictif des difficultés des questions, dans lequel nous définissons les différentes caractéristiques des items, nous illustrons ce modèle avec l'exemple d'un item PISA Science 2015¹. Précisons d'emblée que lorsque nous utilisons le terme général « modèle de difficulté de question », nous entendons dans le cas de l'exemple des items PISA, le modèle de difficulté de l'item comprenant l'introduction et la question de l'item. Enfin nous discutons de la construction de notre modèle, les attentes de résultats qu'il pourrait induire, et l'implication de ce type d'études pour l'enseignement et la conception d'évaluations en sciences.

MODÈLE PRÉDICTIF DE DIFFICULTÉS D'UNE QUESTION

Tout d'abord, il est nécessaire de préciser que la difficulté d'une question est mesurée *a posteriori*, en fonction du taux de réussite des élèves à une question. Plus une question est réussie, plus elle est facile ou encore si le pourcentage de bonnes réponses est bas, la question est considérée comme difficile pour les élèves.

Ces dernières décennies, de nombreuses recherches se sont intéressées à l'identification des facteurs de difficulté d'une question et plus précisément à l'élaboration de modèles prédictifs de difficultés de questions (voir synthèse dans Dhillon, 2011). Il apparaît que les différentes recherches se situent dans des cadres théoriques, issus de la psychologie, des sciences de l'éducation, de la didactique et de la sociologie. Ceci conduit à approcher ces difficultés de diverses manières : d'une part, à partir de cas individuels avec une analyse de l'évaluation passant par la compréhension de ce qu'il se passe lorsque l'élève essaie de répondre à une question et d'autre part, à partir de populations d'élèves et de résultats statistiques de leurs réponses à des items.

Jusqu'à présent, une grande partie des recherches effectuées sur ces modèles ont porté sur des sujets bien précis comme des problèmes d'analogie (par exemple Mulholland *et al.*, 1980; Sternberg, 1977) ou des problèmes algébriques (par exemple Tout *et al.*, 2020; Keehner *et al.*, 2016; Katz *et al.*, 2002) et se sont limitées aux composantes cognitives nécessaires à la résolution. Ainsi les modèles élaborés prenaient en compte différents niveaux de difficulté au sein d'ensembles très homogènes de types d'items. Dans ces études, il est apparu que la difficulté des questions n'était pas seulement prévisible par le nombre et la complexité de chacune des composantes cognitives nécessaires à la résolution mais aussi par les exigences combinées et simultanées qu'elles imposent aux candidats sur le plan cognitif.

D'autres recherches portant par exemple sur des questions de compréhension de lecture (par exemple Pollitt et Ahmed, 1999) ont tenté d'élaborer des modèles plus génériques de la difficulté des questions, avec l'idée de représenter les processus principaux psychologiques impliqués dans la production de réponses aux questions écrites. Plus spécifiquement, trois types de difficultés, ont été distingués (par exemple Pollitt *et al.*, 1985; Ahmed et Pollitt, 1999) :

celui lié au « concept », s'intéressant à la complexité conceptuelle inhérente au savoir et déterminée par la mesure dans laquelle les concepts impliqués dans une question sont abstraits ou concrets; celui lié au « processus » concernant la difficulté des opérations cognitives; celui lié à la « question » faisant référence aux propriétés linguistiques et structurelles de la question.

Dans le même ordre d'idée, Crisp et Grayson (2013) ont proposé une catégorisation tripartite des difficultés de questions dans une évaluation en sciences physiques, comprenant « les attributs de la question », « les processus impliqués dans la question », et « les connaissances et compréhension de la physique ». De même, dans leur analyse secondaire des résultats PISA Science 2000, Prenzel *et al.* (2002) ont proposé trois catégories de sources possibles de difficulté : les caractéristiques formelles de la question (par exemple le format, les illustrations, etc.), les exigences cognitives pour la résolution de l'item (par exemple les différents processus qui ont lieu durant la tâche comme faire des liens logiques ou encore le traitement du texte) et les caractéristiques liées aux connaissances.

Ainsi les recherches qui proposent des modèles de difficulté de questions, en allant au-delà du cognitif strict, ont des points communs. Leurs modèles comportent trois catégories. Il y a unanimité sur la catégorie relative aux caractéristiques intrinsèques à la question, même si les caractéristiques la composant ne sont pas tout à fait identiques. Une autre est relative aux processus cognitifs (traitement du texte, liens logiques). La troisième est relative aux connaissances. Dans notre étude nous reprenons ces trois catégories pour construire notre propre modèle (cf. figure 1). Précisons que ce modèle vise à être utilisé aussi bien dans une perspective d'analyses statistiques reliant les caractéristiques des questions à des profils d'élèves, que dans une pratique évaluative par des enseignants ou des concepteurs de tests.

Des recherches secondaires à PISA se sont concentrées sur l'identification des sources de difficultés chez les élèves de manière quantitative (par exemple Dyrvold, 2020; Pettersen et Braeken, 2019; Le Hebel *et al.*, 2017; Turner *et al.*, 2013) en prenant en compte pour certaines la dimension du SESC des élèves (par exemple Yildirim et Ceylan, 2020). D'autres études secondaires de PISA, ont cherché, du point de vue qualitatif, à observer la manière dont certaines caractéristiques des questions pouvaient

entraîner des variations de performances chez les élèves (par exemple Salgado et Stacey, 2014 ; Le Hebel *et al.*, 2014 ; 2016).

Néanmoins, malgré un certain nombre de recherches portant sur les sources de difficulté des questions, il reste encore difficile d'arriver à un consensus sur l'ensemble des caractéristiques de questions posant difficulté aux élèves (Dong, 2020).

QUESTION DE RECHERCHE

Rappelons que notre but est de proposer un modèle de difficulté de questions de sciences permettant d'anticiper les difficultés des élèves. Cela implique une analyse *a priori* de ces questions. Notre question de recherche est donc la suivante : Quelles caractéristiques des questions évaluant la culture scientifique peuvent conduire à des difficultés chez les élèves ? Nous faisons l'hypothèse que certaines peuvent être spécifiques du milieu socio-économique et culturel et/ou du niveau de performance des élèves.

Nous testerons statistiquement ce modèle dans un travail ultérieur, avec un processus itératif afin de déterminer quelles caractéristiques discriminent plus spécifiquement les élèves selon leur SESC et/ou selon leur niveau de performance.

MÉTHODOLOGIE

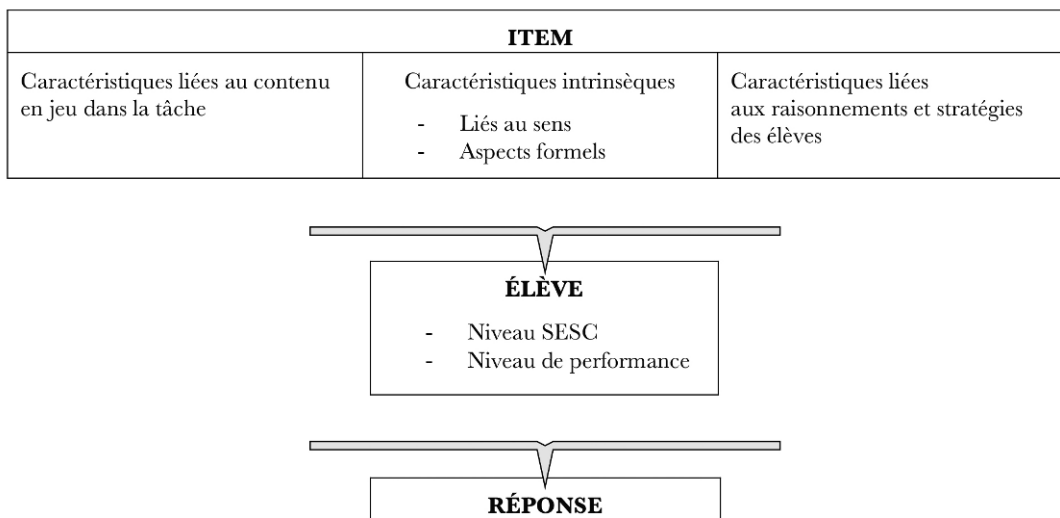
Pour construire un tel modèle, nous nous appuyons sur les travaux présentés ci-dessus. Nous nous servons également de nos travaux antérieurs relatifs à la compréhension de la façon dont les élèves construisent leur réponse aux questions relevant de la culture scientifique (items PISA Science 2006) et aux difficultés qu'ils rencontrent (Le Hebel *et al.*, 2014 ; 2016). Ces travaux correspondent à des études de cas d'élèves filmés lors de la construction de leurs réponses.

L'ensemble de ces travaux nous a permis d'établir trois grandes catégories de caractéristiques potentiellement sources de différents types de difficultés rencontrées par les élèves lors de l'élaboration d'une réponse à une question (cf. figure 1).

L'étude complète nécessite de procéder de manière itérative :

- en construisant un modèle hypothétique comportant des caractéristiques de questions pouvant *a priori* présenter des difficultés pour les élèves, dont certaines spécifiques selon les différentes populations d'élèves (c'est-à-dire selon leurs statuts socio-économico-culturels et leurs niveaux de performance) ;
- en testant ce modèle statistiquement sur l'échantillon français de PISA 2015, ceci permettant de mesurer l'influence d'une caractéristique ou de la conjonction de plusieurs

Figure 1. Schéma pour construire un modèle de difficulté de la question : Interaction entre l'élève et l'item selon les 3 catégories de caractéristiques de l'item dans le processus de réponse



caractéristiques sur l'échec ou la réussite de la question par l'élève. Les résultats de cette étude statistique seront développés dans une publication ultérieure.

Dans nos études antérieures, nous sommes partis du constat empirique de la variété des difficultés rencontrées par les élèves lorsqu'ils résolvent des questions relevant de la culture scientifique (items PISA Science 2006), et ceci nous a conduits à ne pas nous limiter aux difficultés liées spécifiquement au savoir lorsqu'ils construisent leurs réponses (Le Hebel *et al.*, 2014 ; 2016).

Notre modèle *a priori* est appliqué à l'ensemble des 183 items de PISA Science 2015. Il se divise selon les trois catégories suivantes :

La première catégorie est celle des « contenus en jeu » et répertorie quatre caractéristiques. La deuxième concernant les « caractéristiques intrinsèques de l'item » se subdivise en deux sous-catégories : « les aspects formels » et « les aspects liés au sens » comportant l'une et l'autre quatre caractéristiques. Enfin la dernière catégorie « stratégies et raisonnements » se compose aussi de quatre caractéristiques. Cela fait un total de seize caractéristiques principales.

Cela nous a conduits à proposer des sous-caractéristiques pour affiner le pouvoir explicatif de certaines des caractéristiques principales. Ces sous-caractéristiques résultent soit du croisement de deux caractéristiques principales soit de la déclinaison d'une caractéristique principale. Pour une meilleure compréhension des caractéristiques qui composent la dernière catégorie « stratégies et raisonnements », nous exemplifions chacune d'entre elles.

CONSTRUCTION DES CATÉGORIES DE CARACTÉRISTIQUES

Nous présentons l'élaboration des caractéristiques potentiellement sources de différents types de difficultés rencontrées par les élèves pour chacune des trois grandes catégories : les caractéristiques liées au contenu en jeu dans l'item, les caractéristiques intrinsèques de l'item (aspects formels ; aspects liés au sens) et les caractéristiques liées aux raisonnements et stratégies des élèves. Pour fonder ces caractéristiques, nous nous sommes appuyés, selon leur type, sur des travaux et des analyses très différents et c'est ce qui, selon nous, constitue la richesse de ce travail.

L'intégration des caractéristiques à notre modèle de difficulté de question s'est ainsi effectuée en fonction des recherches antérieures. Nous avons dû faire un choix parmi toutes les caractéristiques qui peuvent potentiellement poser problème aux élèves lors de la résolution d'une tâche en sciences. Par exemple, nous avons intégré la longueur du texte à notre modèle mais d'autres caractéristiques linguistiques auraient pu aussi être testées. Toutefois, il était nécessaire de sélectionner un nombre raisonnable de caractéristiques au risque de perdre en significativité statistique. Toutes les caractéristiques et sous-caractéristiques de notre modèle de difficulté des questions en sciences sont détaillées dans cette section et sont présentées dans leur ensemble au sein du tableau 5.

Caractéristiques liées aux contenus en jeu dans la tâche

Pour les caractéristiques liées aux contenus en jeu dans la tâche, nous nous sommes essentiellement basés sur le codage proposé dans le cadre d'évaluation PISA Science 2015 (OCDE, 2016). Dans ce cas, nous précisons les scores obtenus en France, à partir des différentes modalités des caractéristiques, afin de montrer que certaines modalités peuvent être discriminantes sur l'ensemble de l'échantillon français.

La catégorie de caractéristiques « liées aux contenus en jeu dans la tâche » regroupe donc quatre caractéristiques principales (tableau 1) : le type de connaissances, les systèmes, le type de compétences et les champs d'application. Nous proposons aussi une sous-caractéristique qui croise les connaissances de contenu avec les systèmes.

Tableau 1. Caractéristiques et sous-caractéristiques liées aux contenus en jeu dans la tâche

Catégorie	Caractéristique Sous-caractéristique
Caractéristiques liées aux contenus en jeu dans la tâche	1 Type de connaissances*
	2 Systèmes *1 Croisement connaissances de contenu et systèmes
	3 Type de compétences *
	4 Champs d'application*

*Codé par PISA Science 2015.

Types de connaissances

Selon le cadre d'évaluation de PISA, les élèves doivent posséder un « bagage » de connaissances en sciences ainsi que des connaissances *à propos* de la science pour répondre aux différentes questions. Les connaissances *en* sciences aussi appelées « connaissances scientifiques » (nous les nommerons « connaissances de contenu ») sont relatives aux faits, aux concepts et théories établies par la communauté scientifique (OCDE, 2016). Les connaissances *à propos* des sciences (OCDE, 2007) sont scindées en deux sous-catégories dans PISA Science 2015 : les connaissances procédurales relatives aux concepts sur lesquels se fondent les études empiriques et les connaissances épistémiques qui renvoient notamment au fait de comprendre le bien-fondé des pratiques caractérisant la démarche scientifique (OCDE, 2016) (cf. tableau 5).

Les résultats PISA Science 2015 montrent que selon les pays, les scores des items testant les connaissances de contenu sont légèrement inférieurs ou supérieurs aux scores des items testant les connaissances procédurales et épistémiques² (OCDE, 2016, p. 104).

Systèmes

« Les items constituant les épreuves PISA de sciences administrées en 2015 portent sur des sujets empruntés aux disciplines scientifiques majeures que sont la physique, la chimie, la biologie et les sciences de la Terre et de l'Univers » (OCDE, 2016, p. 105). Ces trois disciplines sont nommées « systèmes » dans PISA 2015.

Les résultats PISA Science 2015 montrent que selon les pays, les scores des items correspondant aux trois systèmes sont peu différents, et le sens des écarts varie³.

Sous-caractéristique : croisement des connaissances de contenu et des systèmes

Cette sous-caractéristique résulte du croisement d'une seule des caractéristiques « type de connaissances » (connaissance de contenu) avec les « systèmes » (système physique, vivant, de la Terre et de l'Univers). Ce croisement n'est pas possible avec l'ensemble des connaissances de notre modèle car il impliquerait un nombre trop important de variables ne permettant pas ensuite une comparaison statisti-

quement fiable. De plus, le nombre d'items pour les connaissances épistémiques et procédurales est plus faible par rapport au nombre d'items en lien avec les connaissances de contenu⁴.

Type de compétences

Cette caractéristique renvoie aux trois compétences⁵ scientifiques évaluées dans PISA Science 2015 qui sont les suivantes :

1. Expliquer des phénomènes de manière scientifique
2. Évaluer et concevoir des investigations scientifiques
3. Interpréter des données et des faits de manière scientifique

Les résultats de PISA Science 2015 indiquent que les élèves français ont plus de difficulté avec la première compétence « Expliquer des phénomènes de manière scientifique » par rapport aux deux autres⁶ (sachant que 93 % des items relevant de cette compétence impliquent des connaissances de contenu).

Champs d'application

Les champs d'application de la science de PISA Science 2015 sont : les frontières de la science et de la technologie, la qualité de l'environnement, les ressources naturelles, les risques, la santé et les maladies. Les résultats de PISA Science 2015 indiquent des taux de réussite des élèves français variables suivant les champs d'application des items (le taux de réussite moyen le plus fort se situant dans le champ d'application « risques » et le plus faible dans le champ d'application « santé et maladies⁷ »).

Caractéristiques intrinsèques de l'item – aspects formels

Les aspects formels de la catégorie de caractéristiques « intrinsèques de l'item » regroupent quatre caractéristiques principales : le format de l'item, la longueur du texte, le type d'illustration et la simulation. Ainsi cette catégorie relève d'une vue succincte d'ordre sémiotique et lexical de l'item ; elle n'aborde pas le contenu et sa signification.

Tableau 2. Caractéristiques et sous-caractéristiques liées aux aspects formels de la tâche

Catégorie	Caractéristique Sous-caractéristique
Caractéristiques intrinsèques de l'item – aspects formels	1 Format de l'item*
	2 Longueur du texte
	3 Type d'illustration
	4 Simulation

*Codé par PISA Science 2015.

Format

PISA Science 2015 propose aux élèves de répondre aux questions de 3 façons différentes. Il peut être demandé aux élèves de choisir une réponse parmi d'autres, c'est ce qu'on appelle un choix multiple, celui-ci peut être simple (une seule réponse possible) ou complexe (plusieurs réponses possibles ou plusieurs propositions auxquelles il faut répondre oui/non ou vrai/faux). La question peut aussi donner lieu à une explication de la part de l'élève qui devra alors produire une réponse construite, c'est ce que l'on nomme un format de réponse « ouvert ». Cette mixité des formats de réponse aux questions est assez courante dans les enquêtes internationales (Rauch et Hartig, 2010).

Les Questions à Choix Multiples (QCM) dans PISA⁸ (par exemple Rémond, 2006 ; Le Hebel *et al.*, 2017) ou TIMSS (Liou et Bulut, 2017) sont en général, mieux réussies dans les évaluations par les élèves, en particulier dans les pays francophones.

Cela peut s'expliquer notamment par le fait que différentes réponses soient mises à leur disposition, leur laissant la possibilité de deviner la réponse. De plus, il semblerait que ce genre de format les rassure sur leur capacité à répondre à la question (Simkin et Kuechler, 2005).

Le format réponse ouverte peut sembler plus difficile aux élèves, car ils doivent organiser une réponse et cela leur demande des exigences cognitives plus élevées (Ahmed et Pollitt, 1999). Par exemple, dans une étude secondaire effectuée sur le test national KS2 en Angleterre (El Masri *et al.*, 2017), le format réponse ouverte apparaît comme le prédicteur le plus élevé de la difficulté de l'item dans ce test.

Pour illustrer la différence entre les deux formats de question, Grapin et Sayac (2017) ont comparé les

réponses d'élèves de CM2 (grade 5) à deux tests de transformations d'écriture de nombres aux formats différents (QCM vs Réponse ouverte) mais équivalents du point de vue de la tâche mathématique. L'étude des réponses montre que les élèves ne mobilisent pas les mêmes connaissances d'un format à l'autre. Un changement, du point de vue cognitif, semble aussi se produire, entre le format ouvert qui amène à la production d'une représentation d'un nombre et le QCM qui entraîne plutôt la reconnaissance d'une représentation d'un nombre.

Longueur du texte

Pour Bergqvist et Österholm (2010) les processus de lecture et les processus de résolution de problèmes sont indissociables. De ce fait, il est possible que les caractéristiques linguistiques (comme la longueur du texte) d'une tâche en sciences (notamment les items PISA Science s'appuyant beaucoup sur des textes), influencent sa lecture, sa compréhension et donc sa résolution (Bergqvist *et al.*, 2018 ; Schmitt *et al.*, 2011).

Une des méthodes utilisées pour estimer la longueur du texte dans une tâche est de comptabiliser le nombre total de mots dans celle-ci (par exemple White, 2012). Ainsi plus la tâche contient de mots plus le contenu à traiter est dense. Bergqvist *et al.* (2018) suppose qu'il pourrait être plus difficile de traiter l'ensemble du contenu si la tâche est longue plutôt que courte.

En effet une tâche avec un long texte à lire peut, d'emblée, être perçue par les élèves comme difficile et les démotiver dans leur engagement à la résoudre (Milenković *et al.*, 2015). Une analyse secondaire de PISA 2006 a montré que, pour les élèves en difficulté, un texte long cumulé à une non-familiarité du contexte peut être particulièrement désavantageux car ils identifieront plus difficilement quelles informations sont les plus pertinentes dans le texte (Le Hebel *et al.*, 2017). De plus, sur des tâches de compréhension de l'écrit issues de PISA 2000, les versions françaises sont plus longues que les versions anglaises et s'avèrent ainsi plus difficiles pour les élèves (Grisay, 2003, cité dans Bergqvist *et al.*, 2018, p. 45). Pourtant les études empiriques peuvent différer dans leurs conclusions sur l'effet significatif de la longueur du texte sur la difficulté de la tâche. Ainsi Bergqvist *et al.* (2018) n'ont finalement pas démontré d'effet significatif de la longueur du texte sur la

difficulté d'une tâche en mathématiques et il en va de même en science (Prenzel *et al.*, 2002). Selon les résultats de Stiller *et al.* (2016), c'est plus spécifiquement la longueur des options de réponse qui augmente la difficulté de la tâche. De manière plus générale, Mullis, Martin et Foy (2013) constatent que les élèves d'un niveau bas en mathématiques sont désavantagés par des items (TIMSS 2011) avec une exigence élevée en lecture (estimée selon quatre critères dont le nombre de mots).

Il semblerait donc que cette caractéristique « longueur du texte », que nous mesurons dans notre modèle par la comptabilisation du nombre total de mots dans la tâche (mots du texte et mots présents dans les illustrations), puisse parfois entraîner des variations de performances, c'est pourquoi elle nous paraît importante à étudier dans ce modèle prédictif des difficultés de la question.

Pour mieux caractériser la longueur du texte, nous proposons un découpage des items en quartiles selon le nombre de mots recensés dans chaque item⁹ (quartiles 1 : items de 44 à 136 mots ; quartile 2 : items de 137 à 195 mots ; quartile 3 : items de 196 à 244 mots ; quartile 4 : items de 247 à 417 mots) (cf. tableau 5).

Type d'illustrations

La diversité des représentations est essentielle dans la communication scientifique et dans son enseignement (par exemple Airey et Linder, 2009). Par exemple la construction du sens dans les textes scientifiques passe par les mots mais aussi par une diversité de représentations comme les graphiques, tableaux, diagrammes, cartes, dessins, photographies (McTigue et Flowers, 2011 ; Lemke, 1998). Il en va de même pour les évaluations en science, dont PISA, qui sont constituées de texte et d'illustrations variées (El Masri *et al.*, 2017).

Il nous semble alors important de comprendre la manière dont les élèves exploitent les illustrations dans les items PISA pour mieux envisager quelles illustrations pourraient ou non engendrer de la difficulté.

Les illustrations sont souvent utilisées, dans les manuels notamment, dans le but de motiver davantage les élèves et de les aider à mieux comprendre (Drouin, 1987). Seulement, le traitement d'une illustration du type « image » est relativement rapide (environ 300 millisecondes) et cela peut donner l'illusion aux élèves qu'ils ont parfaitement compris

alors qu'ils ont traité superficiellement ou inadéquatement les illustrations d'une tâche (Peeck, 1993).

Reid et Beveridge (1986) ont d'ailleurs remarqué qu'il y avait des différences d'effet des illustrations en fonction des capacités scientifiques des élèves, elles peuvent ainsi les aider ou plutôt les distraire selon qu'ils ont des capacités élevées ou faibles. Ainsi, pour les élèves qui ont des difficultés à comprendre le texte, le fait qu'il y ait des images intégrées à celui-ci n'augmenterait pas forcément leur performance (Reid et Beveridge, 1986). Pour ces auteurs, le fait que certaines recherches aient démontré un effet positif des images sur la performance des élèves pourrait être dû, en partie, à la non prise en compte des capacités des élèves en sciences.

Hannus et Hyönä (1999) ont utilisé l'*eye tracking* afin de mieux comprendre la manière dont les élèves portaient leur attention sur les illustrations lors de l'apprentissage de passages d'un manuel de biologie, composés de textes et d'illustrations de type « photographies » ou « dessin ». En complément, ils ont mesuré l'intelligence fluide, ou capacité à raisonner avec de nouvelles informations (Rey-Mermet *et al.*, 2019), à l'aide des matrices de Raven, pour estimer les capacités intellectuelles des élèves. Après avoir distingué les segments d'informations pertinents de ceux non pertinents dans les textes et illustrations des documents proposés aux élèves, les auteurs ont conclu qu'en fonction de leurs capacités intellectuelles, les élèves ne prêtent pas attention aux mêmes informations. Les élèves ayant obtenu de meilleurs résultats au test d'intelligence fluide ont tendance à être plus attentifs aux segments catégorisés comme pertinents par les chercheurs, que ce soit dans le texte ou les illustrations.

Ces deux études montrent que certains élèves ont plus de difficultés à traiter efficacement les illustrations dans l'item et que cela peut jouer sur leurs performances car les illustrations sont parfois centrales dans la tâche. Il semble donc intéressant de mieux comprendre l'impact des illustrations dans une situation d'évaluation des sciences, c'est pourquoi nous intégrons cette caractéristique « illustration » à notre modèle de difficulté à l'item (cf. tableau 5).

Simulation

La grande nouveauté de PISA Science 2015 réside dans le fait que la passation du test a été réalisée sur

ordinateur. Cela a ainsi permis d'intégrer des simulations au test et ainsi se rapprocher de l'idée d'investigation scientifique. Dans ces simulations, c'est à l'élève de modifier plusieurs paramètres de la situation puis d'enclencher le système afin de simuler ce qui va se passer (Robinson et Schraw, 2008). Un résultat est donc produit sous l'effet de l'action de l'élève. Ainsworth (2008) a cherché à analyser l'influence que les animations pouvaient avoir sur l'apprentissage. Elle conclut que la majorité des élèves semble davantage motivée à résoudre la tâche de par la présence d'animations. Cependant, ces dernières peuvent entraîner une « illusion de compréhension » lorsque les élèves les traitent de manière passive en observant juste la dynamique proposée (Lowe, 2003, cité dans, Ainsworth, 2008, p. 14).

Il nous semblait donc important d'intégrer cette caractéristique à notre analyse *a priori* afin de mieux juger de son impact sur la difficulté de l'item, d'autant plus que les taux de réussite moyens des items PISA Science 2015 faisant appel à la simulation sont légèrement plus faibles par rapport aux items sans simulation¹⁰.

Caractéristiques intrinsèques de l'item – aspects liés au sens

Les aspects liés au sens de la catégorie de caractéristiques « intrinsèques de l'item » regroupent quatre caractéristiques principales : les contextes, la réponse dans l'item, la dépendance aux ressources, le lien question-illustration. Nous avons trois sous-caractéristiques, deux issues du croisement entre la dépendance et les contextes et la dernière déclinant la présence d'un lien question-illustration selon sa nature (explicite ou implicite). Cette catégorie est donc relative aux caractéristiques de l'item qui entrent en jeu pour la compréhension de la situation présentée en lien avec la question posée.

Contextes

Les items PISA Science 2015 sont tous insérés dans un contexte plus ou moins proche de la vie des élèves. Ils font référence à la vie personnelle ou sociale du citoyen puisque l'objectif de PISA est d'évaluer les connaissances scientifiques des élèves de 15 ans, mais également la manière dont ils

Tableau 3. Caractéristiques et sous-caractéristiques relatives aux aspects liés au sens

Catégorie	Caractéristique Sous-caractéristique
Caractéristiques intrinsèques de l'item – aspects liés au sens	1 Contextes**
	2 Réponse dans l'item
	3 Dépendance-indépendance aux ressources 1 Croisement dépendance aux ressources et contextes 2 Croisement dépendance aux ressources et contextes selon leur forme (simple ou double)
	4 Lien question-illustration 3 Lien explicite/implicite Déclinaison de la présence d'un lien question-illustration selon sa nature (explicite ou implicite)

**Recodage effectué par rapport à celui initialement proposé par PISA Science 2015

les mettent en œuvre dans des situations de la vie courante, en référence à la définition de la culture scientifique.

Comme le précise Wijaya, Van den Heuvel-Panhuizen et Doorman (2015) les items PISA sont d'une part, présentés dans une situation donnée pouvant se rapporter à un contexte réel ou imaginaire et d'autre part, variés en termes de données contextuelles puisqu'ils peuvent contenir des informations relatives à des aspects personnels, scientifiques et publics. PISA Science 2015 distingue trois catégories de contexte : « personnel », « local-national » ou plutôt « mondial » (OCDE, 2016).

La contextualisation des tâches en science consiste à les présenter dans un contexte où la situation exposée fait référence à la vie personnelle ou sociale des élèves. Les effets de l'enseignement contextualisé en chimie ont notamment été étudiés par Ulusoy et Onen (2014) qui ont observé une augmentation des performances des élèves, sur des questions de chimie liées aux expériences quotidiennes. Selon les auteurs, ces meilleures performances découlent d'une motivation plus accrue vis-à-vis de l'apprentissage en contexte de la chimie.

Il semblerait que cette hausse de motivation provoquée chez les élèves par les tâches en contexte quotidien soit due au caractère authentique des

situations évoquées dans la mesure où celui-ci se rapproche de la réalité du rôle endossé par les scientifiques (Roth et Roychoudhury, 1993).

Toutefois, Le Hebel *et al.* (2017) semblent indiquer que l'effet du contexte (positif ou négatif) dépend finalement de la représentation que l'élève s'en fait.

En effet, trois niveaux de représentation d'un texte sont distingués : la structure de surface, la base de texte (idées principales) et le modèle de situation. Il se trouve que pour bien comprendre la situation décrite dans un énoncé, il faut que les élèves arrivent à accéder à un modèle pertinent de situation (correspondant au niveau de représentation le plus élevé) (Van Dijk et Kintsch, 1983). Ce dernier niveau inclut une connexion avec les connaissances et expériences antérieures et implique de réaliser des inférences. Autrement dit, l'interprétation du contexte par l'élève dépend de ses propres expériences (Boaler, 1993) et de ses connaissances antérieures (Verschaffel *et al.*, 2000, cité dans Vicente et Machado, 2016).

L'élève peut donc accéder aux idées principales du texte (base du texte) sans pour autant les utiliser de façon appropriée lors de la résolution de problème (Vidal-Abarca et Sanjose, 1998) s'ils ne les connectent pas aux expériences personnelles adéquates et aux connaissances requises. C'est de cette façon que l'élève peut activer des représentations qui n'ont aucun rapport avec ce qui est attendu de lui au cours de la résolution du problème (Pollitt et Ahmed, 1999), en particulier chez les élèves les moins performants (Wijaya *et al.*, 2014; Le Hebel *et al.*, 2017). Ces derniers sont davantage distraits par le flux d'informations à leur disposition dans le contexte proposé et montrent des difficultés à sélectionner les informations pertinentes pour la résolution de la tâche contrairement aux élèves performants (Le Hebel *et al.*, 2017). Par ailleurs, Hannus et Hyönä (1999) ont montré que les élèves à capacités intellectuelles basses (mesurées à l'aide des matrices de Raven) ont tendance à prêter moins attention aux informations pertinentes du texte de la tâche scientifique contrairement aux élèves présentant de hautes capacités.

En plus de devoir sélectionner les informations pertinentes de celles qui ne le sont pas, il s'agit aussi parfois pour l'élève, sur certains items, de faire une distinction entre les éléments scientifiques pertinents du contexte et les aspects relevant davantage du quotidien afin de traiter la tâche comme un problème scientifique (Serder et Jakobsson, 2016). Distinguer les informations nécessaires pour

répondre à la question peut avoir pour conséquence une potentielle augmentation de la charge cognitive (Milenković *et al.*, 2015; El Masri *et al.*, 2017).

Cela suggère donc qu'on ne peut pas conclure d'un effet uniforme du contexte sur la difficulté des tâches (Boaler, 1993). En effet, même si le contexte semble pouvoir motiver les élèves, il peut aussi poser problème à certains d'entre eux (Dong, 2020); c'est pourquoi il est intéressant d'intégrer cette caractéristique à notre étude afin d'envisager plus précisément l'effet de la contextualisation de la tâche sur les performances des élèves. Ainsi, la classification des items PISA selon trois types de contextes différents ne nous paraissait pas assez précise¹¹. C'est pour cette raison que nous l'avons enrichie en proposant, non pas trois mais cinq types de contextes possibles (avec pour certains une combinaison de deux contextes distincts) : Global¹²; Sociétal; Global-personnel, Sociétal-Global; Sociétal-Personnel (cf. tableau 5).

Réponse dans l'item

Certains items PISA Science 2015 contiennent la réponse soit sous forme textuelle par le biais d'une phrase soit sous forme visuelle via différents symboles présents dans les illustrations qui peuvent indiquer aux élèves où se trouve la réponse. Il faut également noter que les items pour lesquels la réponse est présente au sein des ressources (textes, illustrations) sont majoritairement proposés selon un format à choix multiples. Néanmoins, un seul item au format de réponse ouverte est concerné par cette caractéristique de « réponse dans l'item ».

Les résultats PISA Science 2015 montrent des taux de réussite moyens plus élevés pour les items contenant la réponse (0,6864) que pour les items ne contenant pas (0,4828). On peut faire l'hypothèse qu'un item qui contiendrait la réponse sera moins coûteux à traiter pour les élèves au moment de donner la réponse.

Dépendance-indépendance aux ressources

Les ressources renvoient à la notion utilisée par Ahmed et Pollitt (1999) pour désigner les éléments rattachés à la question comme le texte et les différentes illustrations qui l'accompagnent. Selon eux, l'utilisation des ressources intervient au même

moment que la recherche des connaissances sur le sujet, juste avant de réaliser une correspondance entre la représentation de la question et la connaissance du sujet.

Bien que tous les items PISA Science 2015 soient inclus dans un contexte riche en ressources disponibles avec texte(s) et illustration(s), il n'est parfois pas nécessaire de faire appel à ces ressources pour construire une réponse correcte à l'item, en particulier pour les items qui requièrent une application de connaissances. En effet, pour les items relevant de la sous-compétence « se remémorer les connaissances scientifiques appropriées et les appliquer » (incluse dans la compétence majeure « Expliquer des phénomènes de manière scientifique »), les élèves doivent avoir davantage recours à leurs ressources dites internes, c'est-à-dire aux connaissances dans le domaine scientifique acquises jusqu'à présent.

Ainsi, nous avons distingué les items dont le traitement des ressources est indispensable pour l'élaboration d'une réponse (items dépendants aux ressources) de ceux dont les ressources ne sont pas nécessaires pour répondre correctement (items indépendants aux ressources), ces derniers faisant appel pour la plupart à des connaissances de contenu (84 %).

De plus, pour les items « indépendants aux ressources » nous supposons que le contexte est davantage mis au second plan par les élèves, contrairement aux « items dépendants » aux ressources. Cela nous a donc amené à compléter notre catégorisation avec deux sous-caractéristiques.

Sous-caractéristique : *croisement de la dépendance aux ressources avec les contextes*

Cette sous-caractéristique résulte du croisement de la caractéristique « dépendance-indépendance aux ressources » et de celle des « contextes » (Global, Sociétal, Global-personnel, Sociétal-Global, Sociétal-Personnel). Nous souhaitons comparer l'impact du type de contexte présent dans les items « dépendants aux ressources » sur les performances des élèves par rapport à ceux dont le contexte est au « second plan » (items indépendants aux ressources).

Sous-caractéristique : *croisement de la dépendance aux ressources et des contextes selon leur forme : simple ou double*

Cette sous-caractéristique résulte du croisement de la caractéristique « dépendance-indépendance aux

ressources » et de celle des « contextes » selon leur forme « simple ou double ». Dans les cinq types de contexte Global, Sociétal, Global-personnel, Sociétal-Global, Sociétal-Personnel, les deux premiers sont considérés comme simples tandis que les trois autres comme doubles. Nous souhaitons comparer le contexte simple ou double des items « dépendants aux ressources » à ceux dont le contexte est au « second plan » (items indépendants aux ressources).

Présence d'un lien entre la question et l'illustration

Plus des 3/4 des 183 items PISA Science 2015 contiennent une illustration (par exemple graphique, photo-dessin, schéma, tableau, multiples illustrations). Cependant le lien entre la question et l'illustration n'est pas toujours existant.

En effet, certaines illustrations, considérées comme représentatives, font seulement écho à une partie du texte (Choplin *et al.*, 2001 ; Carney et Levin, 2002). Par exemple dans PISA Science, les illustrations contenues dans les items peuvent renvoyer, mais pas de manière systématique, à des éléments évoqués dans le texte introductif de l'unité¹³ PISA et/ou de la question de l'item. Autrement dit, dans PISA Science 2015, l'illustration n'est pas systématiquement utile pour donner une réponse à la question.

Le lien entre la question et l'illustration n'est pas à confondre avec la catégorie précédente « dépendance de la question aux ressources disponibles ». En effet, dans notre codage des items « dépendants aux ressources », nous considérons le texte et l'illustration comme un tout alors que dans le « lien question-illustration » nous les considérons comme deux entités différentes. Autrement dit, le codage de cette caractéristique est basé seulement sur l'utilité ou non de l'illustration pour répondre à la question.

La caractéristique « lien question-illustration » nous permet ainsi de prendre en compte le rôle joué plus spécifiquement par l'illustration (indépendamment du texte) dans le processus de réponse à la question.

Sous-caractéristique : *déclinaison de la présence d'un lien question-illustration selon la nature du lien (explicite ou implicite)*

Néanmoins, malgré la présence d'un lien entre la question et l'illustration, celui-ci peut être parfois

explicitement donné au cours d'une instruction écrite dans l'item ou bien devoir être déduit par l'élève (implicitement). Nous supposons que des indices donnés dans le texte en rapport avec les illustrations permettent de guider les élèves sur la façon d'associer les deux sources d'information et leur indiquent également le moment propice au traitement des illustrations (Slough *et al.*, 2010).

Il nous semble donc important de spécifier l'effet de cette différence d'explicitation des consignes sur la profondeur du traitement réalisé par les élèves sur les illustrations et son influence sur leurs performances. Nous avons donc complété notre catégorisation avec la sous-caractéristique « Lien question-illustration explicite ou implicite ». Elle résulte de la déclinaison de la caractéristique « Lien question-illustration » (lien existant) selon sa forme « explicite ou implicite ».

Caractéristiques liées aux raisonnements et stratégies de réponse des élèves

Un comportement adapté face à la tâche peut être entravé à cause des caractéristiques formelles de l'item, qui peuvent le rendre plus ou moins difficile, mais également à cause de la procédure adoptée durant la résolution de la tâche par l'élève. Ce choix dépend de sa façon de traiter l'information, ses représentations de la question et de la situation, puis les processus et les stratégies qu'il convoque (Vianin, 2009). En fonction de la procédure que l'élève met en place, celle-ci peut l'amener vers la direction souhaitée tandis que parfois elle peut le diriger vers un autre but de la question.

Les caractéristiques liées aux raisonnements et stratégies de réponses des élèves regroupent quatre caractéristiques principales : la complexité cognitive, le matching, la projection et le niveau de référence à la vie quotidienne. Certaines de ces caractéristiques principales ont été complétées par des sous-caractéristiques. Nous avons décliné la présence de matching selon son effet (aidant ou invalidant) et la présence de projection selon le type et la forme (directe ou indirecte).

Ces caractéristiques sont très différentes du point de vue de la spécificité des raisonnements ou stratégies de réponses. Alors que la complexité cognitive comme nous le présentons ci-dessous incorpore l'ensemble de la construction de la réponse, les autres sont spécifiques d'une stratégie (le matching) ou d'un

lien avec le contexte ou la question (projection et niveau de référence à la vie quotidienne).

Le choix de ces caractéristiques, qui ne sont pas spécifiques des sciences, mais générales, est orienté par notre but de discriminer des caractéristiques de questions selon le niveau socio-économique et culturel et la performance.

Nous avons choisi d'exemplifier plus largement ces caractéristiques liées aux raisonnements et stratégies de réponse des élèves car elles sont moins couramment étudiées dans la littérature comme source potentielle de difficulté de questions de sciences. Il nous semblait donc primordial de les illustrer davantage grâce à des questions issues de sources diverses (évaluations internationales, manuels scolaires, etc.).

Par le biais de ces différentes sources d'exemples, nous souhaitons également montrer que notre modèle peut être appliqué dans différents contextes, par exemple, pour un exercice issu d'un manuel scolaire, ou pour différents supports d'évaluations, etc.

Tableau 4. Caractéristiques et sous-caractéristiques liées aux raisonnements et stratégies de réponses des élèves

Catégorie	Caractéristique Sous-caractéristique
Caractéristiques liées aux raisonnements et stratégies de réponse des élèves	1 Complexité cognitive (DOK)**
	2 Matching 1 Matching aidant/invalidant Déclinaison de la présence de matching dans l'item selon son effet (aidant ou invalidant)
	3 Projection 2 Type de projection Déclinaison de la projection présente dans l'item selon le type de projection 3 Projection directe/indirecte Déclinaison de la projection présente dans l'item selon sa forme (directe ou indirecte)
	4 Niveau de référence à la vie quotidienne

**Recodage effectué par rapport à celui initialement proposé par PISA Science 2015.

Complexité cognitive

On peut dire que la complexité cognitive est estimée *a priori* dans PISA Science 2015 et plusieurs taxonomies ou échelles ont déjà été proposées (par

exemple Bloom, 1956 ; Marzano et Kendall, 2007 ; Ford et Wargo, 2012).

Ahmed et Pollitt (1999), par exemple, ont développé une échelle à 4 dimensions afin de mesurer les exigences cognitives d'une tâche : complexité, abstraction, ressources et stratégie (échelle CRAS). La complexité cognitive de la tâche s'estime alors par rapport au nombre d'opérations cognitives à effectuer lors de sa résolution.

Webb (2007) propose une échelle de complexité cognitive à 4 niveaux appelée Depth of Knowledge (DOK) afin d'analyser les exigences cognitives attendues dans le cadre d'évaluations. Les deux premiers niveaux de l'échelle sont relativement peu coûteux en termes d'exigences cognitives : on passe d'un simple rappel d'un élément de connaissance ou de procédure simple (niveau 1) à la compréhension en profondeur d'un seul concept (niveau 2). Pour les deux autres niveaux, les exigences cognitives requises sont plus poussées puisque le niveau 3 (réflexion stratégique) nécessite une compréhension approfondie, avec la planification, l'utilisation de preuves et un raisonnement cognitif plus exigeant. Au niveau 4 (réflexion approfondie), les élèves sont censés établir des liens, relier des idées dans le contenu ou entre domaines de contenu connexes et doivent choisir ou concevoir une approche parmi de nombreuses alternatives à la façon dont la situation peut être résolue.

Dans leur analyse secondaire des items PISA Science 2006, Le Hebel *et al.* (2017) ont montré qu'un item PISA Science de DOK élevé est souvent associé à un fort niveau de difficulté pour les élèves (score de réussite faible), alors qu'un item de DOK faible ou moyen peut avoir des niveaux de difficulté variés. En d'autres termes, les items difficiles (auxquels les élèves ont obtenu des scores faibles) ne relèvent pas forcément d'une complexité cognitive élevée, à l'inverse un niveau de complexité élevé implique leur difficulté.

Pour notre recherche, nous nous sommes basés uniquement sur l'échelle « Depth of Knowledge » (DOK) élaborée par Webb (2007). Nous avons privilégié cette échelle par rapport aux autres car elle a été spécifiquement créée pour estimer la complexité cognitive des items lorsqu'ils sont proposés aux élèves dans des situations d'évaluation des sciences. Elle nous paraît donc adéquate dans le cadre de PISA qui est justement un test évaluatif.

De plus, cette échelle permet de rendre compte de la profondeur des réflexions dont les élèves doivent

faire preuve durant la tâche de résolution. Cette indication est primordiale pour nous permettre d'envisager le raisonnement dans lequel l'élève va devoir s'engager afin de réaliser la tâche.

Dans leur analyse secondaire des items PISA Science 2006, Le Hebel *et al.* (2017) ayant montré que statistiquement le DOK avait une influence sur le niveau de difficulté de l'item, nous décidons d'intégrer cette caractéristique à notre modèle de difficulté à l'item afin d'apporter des données supplémentaires quant à son impact sur les performances selon les différents groupes d'élèves de l'échantillon français 2015.

Un des items libéré PISA Science 2015 nommé « Étude des versants d'une vallée » (question 2/2)¹⁴ est un exemple approprié d'illustration d'une complexité cognitive élevée.

Cet item présente tout d'abord un schéma représentant deux versants d'une vallée sur lesquels sont placés des capteurs de rayonnement solaire, d'humidité du sol et de précipitations. Un tableau rassemble les moyennes des mesures relevées sur les différents capteurs et leurs incertitudes. Deux élèves ne sont pas d'accord sur la raison pour laquelle il y a une différence d'humidité du sol entre les deux versants. Le premier l'explique en indiquant une différence de rayonnement solaire entre les deux versants, et le deuxième justifie sa réponse par l'absence de différence en termes de précipitations. À partir des données disponibles, l'élève doit choisir celui qui a raison.

Pour répondre à cet item, l'élève doit prendre en compte un certain nombre de données de format divers ; un schéma puis un tableau. L'élève doit entreprendre une lecture des données du tableau impliquant une comparaison de résultats de nature différente et des incertitudes associées. Ainsi, cet item nécessite de prendre en compte plusieurs variables et d'établir des liens entre elles. Tous ces éléments renvoient au niveau 4 de l'échelle de DOK de Webb (2007) selon notre codage, correspondant également au dernier niveau de complexité cognitive de la catégorisation proposée par PISA.

Stratégie de « matching » ou « correspondance de mots »

Lorsque les élèves ne comprennent pas une question ou ne connaissent pas la réponse, ils peuvent avoir recours à diverses stratégies pour, malgré tout, tenter de donner une réponse (Dempster et Reddy,

2007). En particulier, les élèves ont recours à une stratégie qui consiste à rechercher des termes communs dans la question et dans les réponses proposées (Dempster, 2006), cela leur permet de sélectionner une des réponses proposées dans le choix multiple en optant pour celle qui évoque un terme également contenu dans la question (Dempster, 2007).

Sous-caractéristique : déclinaison de la présence de matching selon son effet aidant ou invalidant

Cette stratégie nommée « matching » ou correspondance de mots s'observe plutôt dans les items à choix multiples et peut s'avérer soit aidante dans le sens où elle permettra à l'élève de sélectionner la bonne réponse, soit invalidante lorsqu'elle amène l'élève à choisir une réponse erronée (Le Hebel *et al.*, 2016). Nous avons donc créé la sous-caractéristique « matching aidant ou invalidant ». Elle résulte d'une déclinaison de la caractéristique « Stratégie du matching » (présence de cette stratégie) selon son effet « aidant ou invalidant ».

Il semblerait que le caractère invalidant de cette stratégie soit en lien avec le niveau de compréhension des élèves. En effet, Cerdán, Gilabert et Vidal-Abarca (2011) ont tenté d'analyser la manière dont les élèves utilisaient la stratégie de matching en fonction de leur niveau de compréhension. Pour cela, ils ont sélectionné douze questions du test PISA de compréhension écrite dont ils avaient préalablement vérifié qu'aucun mot ne se répétait entre la question de l'item et le texte introductif. En parallèle, ils ont proposé 12 autres questions dans lesquelles ils ont fait apparaître des mots-clés dans des zones distrayantes du texte (c'est-à-dire des zones non liées à la réponse attendue) et qui apparaissaient aussi dans le libellé de la question. Leurs résultats montrent que, pour répondre, les élèves « faibles compreneurs » sélectionnent davantage les informations dans les zones distrayantes du texte, celles au sein desquelles se trouvaient les mots-clés (répétés dans la question).

Les « faibles compreneurs » auraient donc tendance à prêter davantage attention aux indices trompeurs dans l'item. Cela indique leur plus grande disposition à utiliser une stratégie de matching non adéquate. Dans le cadre d'études de cas d'élèves filmés lors de la construction de leurs réponses à des items PISA Science, Le Hebel *et al.* (2016) ont identifié l'utilisation de cette stratégie chez les élèves en difficulté.

Le matching est donc intégré à notre modèle de difficulté de la question (cf. tableau 5). Cela peut

apporter une difficulté supplémentaire à l'item puisque certains mots présents à la fois dans l'énoncé de l'item et dans les choix de réponse peuvent activer la mise en place de cette stratégie chez les élèves. De plus, elle semble être utilisée de manière différente en fonction des élèves et peut donc avoir des répercussions différentes d'un groupe à l'autre sur leurs performances.

La dernière évaluation de TIMSS en 2019, qui consiste à contrôler les connaissances d'élèves de 4^e (grade 8) en sciences de la vie et de la Terre, contient un item pouvant présenter un matching invalidant. Cet item (cf. annexe 1) porte sur le cycle de l'eau dans l'écosystème d'une forêt et est composé d'un schéma représentant la rive d'un lac bordée par une forêt sur laquelle tombe de la pluie en provenance d'un ciel nuageux. On trouve également des flèches à la cime des arbres pointées vers le haut et d'autres vers le bas en dessous du ciel nuageux. Il est alors demandé aux élèves de répondre à la question suivante : « Quel processus du cycle de l'eau est présenté par les flèches qui pointent vers le haut ? ».

L'énoncé de l'item explique qu'« une certaine quantité d'eau de pluie est absorbée par les arbres à partir du sol » et la première réponse du choix multiple commence par « l'absorption de l'eau par les racines des arbres ». On observe que 25,5 % des élèves ont choisi cette réponse A. Nous faisons l'hypothèse qu'une certaine partie des élèves s'est laissé influencer par cette correspondance de mots entre la question et la réponse, les conduisant à une réponse incorrecte (matching invalidant), la bonne réponse étant le choix C « l'évaporation de l'eau des feuilles des arbres dans l'air ».

Projection (présence ou non dans l'item)

Notre analyse des énoncés met en évidence que la situation dans plusieurs items est rédigée de telle sorte qu'elle peut inciter les élèves à se projeter dans une communauté plus ou moins proche d'eux-mêmes pour comprendre la situation présentée dans l'item et élaborer leur réponse. Nous supposons que la réponse est plus atteignable si l'élève réalise la projection suscitée par le contexte de l'item.

La projection peut être différente d'un item à l'autre, c'est pourquoi nous avons proposé deux sous-caractéristiques qui permettent de préciser de quel type est la projection (communauté proche

de l'élève, communauté de savoir-faire, c'est-à-dire projection vers des professions ne représentant pas une autorité scientifique, communauté de savoirs, c'est-à-dire projection vers des professions représentant une autorité scientifique ou projection double et sa forme (directe ou indirecte).

Ces deux sous-caractéristiques « type de projection » et « projection directe/indirecte » résultent d'une déclinaison de la caractéristique « projection » (présence d'une projection dans l'item) selon le type précis de projection et sa forme.

Sous-caractéristique : déclinaison de la présence de projection dans l'item selon le type projection

Certains items suggèrent à l'élève de se projeter vers une communauté de savoirs (professions représentant une autorité scientifique) ou bien une communauté de savoir-faire (professions ne représentant pas une autorité scientifique) ou encore une communauté proche de lui-même comme réaliser une projection en tant qu'élève dans une situation donnée. Parfois deux communautés (projection double) peuvent apparaître dans un même contexte et ce sera à l'élève de réaliser la projection attendue pour répondre à la question tout en inhibant l'autre.

Quand on s'intéresse à l'idée de communauté, on note que chacune développe son propre « contrat de communication » (Charaudeau, 1993, cité dans Rebière, 2013, p. 8). On peut donc émettre l'hypothèse qu'il sera plus facile pour l'élève de comprendre le « contrat de communication » d'une communauté dont il est proche que d'une communauté plus éloignée.

Par exemple, la communauté scientifique communique via un langage particulier et partagé au sein de celle-ci. Il peut donc être plus difficile pour l'élève d'accéder à ce langage puisque ce n'est pas son langage habituel. Cela requiert pour l'élève de modifier sa position énonciative (Rebière, 2013). En effet, même si l'élève étudie cette discipline à l'école, cela ne consiste pas seulement pour lui, à acquérir des connaissances, il s'agit également d'apprendre à agir-penser-parler d'une manière analogue aux spécialistes de la discipline.

L'élève devra donc dépasser son statut d'élève et ne plus penser selon son propre point de vue mais arriver à une réflexion partagée selon les principes et connaissances véhiculées par la communauté (Rebière, 2013).

Cependant le caractère technique de disciplines comme les sciences expérimentales peut laisser

penser qu'elles seraient bien plus complexes que d'autres disciplines. Cela peut semer le doute chez certains élèves et les amener à se demander si cette discipline leur est accessible. L'expertise des scientifiques leur donne une apparence surhumaine aux yeux de certains. En effet, la science tend à s'éloigner du sens commun pour faire preuve d'une objectivité toujours plus rigoureuse. Tout cela peut créer chez les élèves une perte de confiance en leurs jugements et opinions « ordinaires » (Lemke, 1990). Il pourra donc être plus difficile pour certains de dépasser leur statut d'élève pour se projeter vers une autre communauté et avoir une réflexion en conséquence car l'expertise des scientifiques, synonyme d'objectivité, peut créer une distance avec les élèves qui peut les bloquer (Lemke, 1990).

Sous-caractéristique : déclinaison de la présence de projection dans l'item selon sa forme (directe ou indirecte)

Un autre élément à noter est que les communautés vers lesquelles l'élève devra se projeter dans des items PISA peuvent être mentionnées explicitement au cours de la question ou bien être évoquées au sein du contexte sans pour autant être convoquées à nouveau durant la question. Dans ce cas, la projection est donnée de manière indirecte dans l'item et représente donc une information implicite. Nous avons donc complété notre catégorisation avec la sous-caractéristique « Projection directe ou indirecte ». Elle résulte de la déclinaison de la caractéristique « Projection » (présence dans l'item) selon sa forme « directe ou indirecte ».

Nous avons choisi d'inclure cette caractéristique « projection » à notre modèle de difficulté à l'item (cf. tableau 5) car, elle est peu étudiée dans la littérature et représente donc une caractéristique innovante. De plus, nous faisons l'hypothèse qu'elle peut être une source de difficulté variable suivant le SESC des élèves.

Les deux items de l'unité libérée de PISA Science 2006 nommée « Cultures génétiquement modifiées » (cf. annexe 2) illustrent bien le cas d'une projection directe vers une communauté de professions représentant une autorité scientifique. L'introduction de l'unité présente les détails d'une étude scientifique sur les OGM. Le premier item porte sur les facteurs que les scientifiques ont volontairement fait varier dans leur étude. Il s'agit alors pour l'élève de répondre à la question suivante : « Dans l'étude scientifique mentionnée dans l'article, quels sont les

facteurs qu'on a volontairement fait varier? » et de choisir pour deux facteurs proposés si oui ou non c'est le cas. Le deuxième item se base sur une phrase de l'introduction qui dit « on a semé du maïs dans 200 champs à travers le pays » et consiste à demander aux élèves la raison pour laquelle les scientifiques ont utilisé plusieurs sites dans leur étude.

Les deux items incitent donc explicitement l'élève à se projeter vers une communauté de professions représentant une autorité scientifique, c'est-à-dire se mettre à la place des scientifiques de l'étude pour réfléchir selon une démarche scientifique et comprendre la méthodologie expérimentale.

Nous avons également constaté que cette incitation à se projeter vers une communauté de professions représentant une autorité scientifique se retrouve dans les manuels scolaires comme l'illustre l'exercice « comprendre le rôle de la pectine chez les plantes » (cf. annexe 3) issu d'un manuel scolaire collaboratif de SVT pour les élèves de seconde. En effet, l'énoncé de cet item révèle que « les chercheurs ont identifié deux mutants, etc. » ce qui encourage les élèves à se mettre à la place des chercheurs qui ont obtenu ces résultats. Toutefois la projection est indirecte car elle n'est pas explicitement réactivée dans la question.

Référence au quotidien

Pour rappel, les items PISA sont tous intégrés à des situations de la vie quotidienne ou citoyenne et nous observons que le niveau de référence au quotidien peut varier d'un item à l'autre. Nous proposons donc un codage de ces références au quotidien sur trois niveaux : faible, modéré, fort. Ce codage est inspiré de la catégorisation de Piel et Schuchart (2014) réalisée à partir des items de l'étude internationale TIMSS. Nous avons alors adapté cette grille aux items PISA.

Les travaux de Bautier et Rochex (1997) indiquent que les élèves en difficulté n'arrivent pas souvent à discriminer le « moi » de leur expérience quotidienne et celui de l'expérience que l'école leur apporte. Ainsi, il peut être particulièrement difficile pour ces élèves d'envisager la réponse que l'on attend d'eux, lorsque l'on mêle dans des tâches, le registre de l'opinion fondée sur l'expérience quotidienne à celui des savoirs propres à l'expérience scolaire, comme cela peut être le cas dans les tâches PISA présentées en contexte.

D'ailleurs d'après Boaler (1993) ce qui rend complexe les problèmes en mathématiques intégrés à des contextes de « monde réel » est le fait qu'ils puissent faire écho à la subjectivité et l'expérience des élèves. Toutefois, Vicente et Manchado (2016) montrent que la résolution d'un problème difficile est plus simple s'il est « authentique » (qui décrit une situation ressemblant à la vie quotidienne des élèves) plutôt que s'il est standard, et ce d'autant plus lorsque les élèves ont un haut niveau en mathématiques. Ces élèves bénéficient davantage de l'aide que peuvent leur apporter les informations de la situation pour résoudre le problème.

Nous faisons donc l'hypothèse que cette référence au quotidien dans les items PISA Science 2015 peut être source de difficulté pour certains élèves, en particulier selon leur environnement socio-économico-culturel, les conduisant vers l'élaboration d'une réponse incorrecte.

Le troisième item de l'unité libérée PISA Science 2015 nommée « Météoroïdes et cratères¹⁵ » est un exemple de référence au quotidien élevée. Cet item est composé d'une image de météoroïde et d'un schéma représentant trois cratères de diamètre différent. Dans cet item, il est simplement demandé à l'élève de classer les cratères selon la taille du météoroïde qui les a formés, du plus grand au plus petit.

Même si l'observation de cratères de météoroïdes est éloignée de la vie quotidienne des élèves, nous pouvons considérer que la notion de formation d'un trou provoquée par un objet familier (par exemple billes, balles) qui tombe, par exemple dans le sable, est une situation quotidienne que les élèves ont déjà pu vivre durant leur enfance. L'estimation de la taille d'un objet grâce à la trace laissée suite à son impact au sol est donc une situation qui a déjà pu être expérimentée par les élèves, c'est pourquoi nous avons classé cet item comme faisant hautement référence à la vie quotidienne.

Synthèse de l'interaction entre les différentes caractéristiques de l'item et l'élève dans le processus de réponse

Le tableau 5 présente et donne une description de l'ensemble des caractéristiques prises en compte dans notre modèle de difficulté de questions.

La figure 2 représente l'interaction entre les 23 caractéristiques de l'item catégorisées par trois

Tableau 5. Ensemble des caractéristiques et sous-caractéristiques intégrées à notre modèle de difficulté de questions pour les items PISA Science 2015

Catégorie	Caractéristique Sous-caractéristique	Description
Caractéristiques liées aux contenus en jeu dans la tâche	Type de connaissances*	Connaissances associées à la culture scientifique, qui sont en même temps liées et distinctes; il s'agit des connaissances de contenu, épistémiques et procédurales (terminologie de l'enquête PISA Science 2015)
	Systèmes*	Domaine scientifique majeur évalué dans les questions : Systèmes physiques, systèmes vivants et systèmes de la Terre et de l'Univers
	<i>Croisement connaissances de contenu et systèmes</i>	Comparaison des connaissances épistémiques et procédurales aux systèmes des connaissances de contenu
	Type de compétences*	Indicateur PISA correspondant à la capacité des élèves de 15 ans d'utiliser leurs connaissances scientifiques pour identifier des problématiques, acquérir de nouvelles connaissances etc. Trois types de compétences sont évaluées : évaluer des phénomènes de manière scientifique; évaluer et concevoir des investigations scientifiques; interpréter des données et des faits de manière scientifique
	Champs d'application*	Champ d'application des questions (Frontières de la science et de la technologie/ Qualité de l'environnement/ Ressources naturelles/ Risques/ Santé et maladies)
Caractéristiques intrinsèques de l'item – aspects formels	Format de l'item*	Type de réponse aux questions : choix multiple simple ou complexe et réponse ouverte construite
	Longueur du texte	Nombre de mots contenus dans chaque tâche (quartile 1 : 44 à 136 mots, quartile 2 : 137 à 195 mots, quartile 3 : 196 à 244 mots et quartile 4 : 247 à 417 mots)
	Type d'illustration	Présence ou non d'une illustration
	Simulation	Présence ou non d'une simulation
Caractéristiques intrinsèques de l'item – aspects liés au sens	Contextes**	Aspects situationnels de la question (Global, Global-personnel, Sociétal, Sociétal-Global, Sociétal-Personnel)
	Réponse dans l'item	Présence ou non de la réponse dans l'item dans le texte et/ou l'illustration
	Dépendance-Indépendance aux ressources	Nécessité ou non de faire appel aux ressources contenues dans l'item pour construire une réponse correcte
	<i>Croisement dépendance aux ressources et contextes</i>	Prise en considération des contextes (caractéristique contexte) seulement pour les items dépendants des ressources disponibles
	<i>Croisement dépendance aux ressources et contextes selon leur forme (simple ou double)</i>	Présence d'un seul contexte ou deux contextes dans les items dépendant des ressources disponibles
	Lien question-illustration	Présence ou non d'un lien question-illustration qui renvoie au fait que la réponse à la question dépend des informations contenues dans l'illustration présente dans l'item. À l'inverse l'absence de lien question-illustration correspond au fait que soit aucune illustration n'est présente dans l'item soit que l'illustration représentée dans l'item est peu utile pour le choix de réponse
	<i>Lien explicite/implicite-Déclinaison de la présence d'un lien question-illustration selon sa nature (explicite ou implicite)</i>	Référence explicite dans la consigne ou la question à la nécessité d'utiliser l'illustration pour répondre à la question ou cette nécessité peut ne pas être mentionnée alors que l'illustration est utile pour répondre à la question (implicite)

Caractéristiques liées aux raisonnements et stratégies de réponse des élèves	Complexité cognitive (DOK)**	Échelle de complexité cognitive à 4 niveaux : niveau 1 (remémoration), niveau 2 (utilisation de compétences et/ou de connaissances conceptuelles), niveau 3 (réflexion stratégique) et niveau 4 (réflexion étendue)
	Matching	Présence ou non d'une stratégie possible de « matching » (correspondance de mots dans la construction de la réponse) entre l'introduction de l'unité/item, la question et les choix de réponses
	<i>Matching aidant/invalidant</i> <i>Déclinaison de la présence de matching dans l'item selon son effet (aidant ou invalidant)</i>	Dans le cas où l'item peut potentiellement conduire à une stratégie de « matching », efficacité de la stratégie si elle aide à trouver la bonne réponse (matching aidant) ou bien invalidante si elle incite plutôt l'élève à se diriger vers une réponse erronée (matching invalidant)
	Projection	Évocation d'une projection dans l'item ou non
	<i>Type de projection</i> <i>Déclinaison de la projection présente dans l'item selon le type de projection</i>	Incitation dans le contexte de la question des élèves à se projeter et à concevoir le point de vue d'une communauté plus ou moins proche de leur propre vie (communauté proche, de savoir-faire, de savoirs -professions représentant une autorité scientifique- ou projection double)
	<i>Projection directe/indirecte</i> <i>Déclinaison de la projection présente dans l'item selon sa forme (directe ou indirecte)</i>	Dans le cas d'une évocation d'une projection dans l'item, soit la projection à réaliser est explicitement mentionnée (directe), la projection est implicite dans la question (indirecte)
	Niveau de référence à la vie quotidienne	Échelle pour rendre compte du niveau d'intensité des références au quotidien dans l'item. Plus la référence au quotidien est forte plus l'item offre à l'élève la possibilité d'une part, de se référer à des objets familiers ou à ses expériences personnelles et d'autre part, de se projeter dans la situation de l'item, au travers de sa sensibilité propre

*Codé par PISA Science 2015 **Recodage effectué par rapport à celui initialement proposé par PISA Science 2015.

des auteurs de l'article, sur l'ensemble des 183 items Science PISA 2015, et l'élève dans le processus de réponse. L'analyse *a priori* compte 16 caractéristiques principales (dont 7 étaient préétablies par le cadre d'évaluation de PISA Science 2015) et 7 sous-caractéristiques correspondant aux déclinaisons de certaines caractéristiques principales.

MISE EN APPLICATION DE LA CATÉGORISATION PROPOSÉE À PARTIR D'UN ITEM PISA SCIENCE 2015 LIBÉRÉ

Nous proposons à présent d'illustrer notre catégorisation, établie dans l'analyse *a priori*, avec un item PISA Science 2015 libéré. Il s'agit du 2^e item de l'unité intitulée « ferme aquatique durable » (figure 3). Nous allons détailler pour cet item chaque choix établi pour les 16 caractéristiques et 7 sous-caractéristiques (tableau 6).

Cet item libéré débute par une page d'introduction générale qui présente la raison pour laquelle la créa-

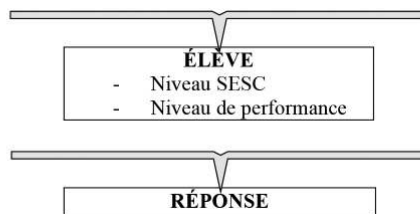
tion d'une ferme aquatique durable est nécessaire et les défis que pose celle-ci. L'introduction se termine par une phrase potentiellement importante pour répondre à l'item ci-dessus : « Les déchets et l'excès de nutriments (la nourriture dont les algues et les plantes ont besoin pour leur croissance) sont éliminés de l'eau avant qu'elle ne retourne dans l'océan ».

Cet item, se caractérise par plusieurs éléments (tableau 6) notamment par les différents contenus mis en jeu au sein de celui-ci. On peut ainsi voir que la connaissance évaluée est de type « contenu » et fait référence au domaine des « systèmes vivants » avec pour sujet spécifique la qualité de l'environnement. La compétence requise pour cet item est « Interpréter des données et des faits de manière scientifique ».

Concernant les caractéristiques intrinsèques de l'item, en termes d'aspects formels, c'est un item à choix multiple simple puisqu'il requiert la sélection d'une seule réponse mais il comporte un texte à lire relativement long (238 mots, quartile 3) et un schéma à comprendre. On peut voir, en termes d'aspects davantage liés au sens, que l'item présente

Figure 2 : Modèle de difficulté de la question détaillé :
Interaction entre les différentes caractéristiques de l'item et l'élève dans le processus de réponse

ITEM		
Caractéristiques liées au contenu en jeu dans la tâche	Caractéristiques intrinsèques de l'item	Caractéristiques liées au raisonnements et stratégies des élèves
1/ Type de connaissances 2/ Systèmes - Croisement connaissances de contenu et systèmes 3/ Type de compétences 4/ Champs d'application	Aspects formels 5/ Format 6/ Longueur du texte 7/ Type d'illustration 8/ Simulation Aspects liés au sens 9/ Contextes 10/ Réponse dans l'item 11/ Dépendance-indépendance aux ressources - Croisement dépendance aux ressources et contextes - Croisement dépendance aux ressources et contextes selon leur forme (simple ou double) 12/ Lien question-illustration - Lien explicite/implicite Déclinaison de la présence d'un lien question-illustration selon la nature du lien (explicite ou implicite)	13/ Complexité cognitive (DOK) 14/ Matching - Matching aidant/invalidant Déclinaison de présence de matching selon son effet (aidant ou invalidant) 15/ Projection - Type de projection Déclinaison de la projection présente dans l'item selon le type de projection - Projection directe/indirecte Déclinaison de la projection présente dans l'item selon sa forme (directe ou indirecte) 16/ Niveau de référence à la vie quotidienne



un lien avec l'illustration car certaines informations renseignées (par exemple la légende « l'eau est nettoyée dans ce bassin » sous le dernier bassin, au sein duquel les herbes de marais sont censées se trouver) dans le schéma peuvent être utiles pour répondre. Toutefois ce lien est implicite puisqu'à aucun moment il n'est précisé qu'il est utile de se référer à l'illustration pour répondre à la question.

La réponse attendue par PISA pour cet item est la réponse 4 « herbes des marais » et on peut voir que la réponse est textuellement donnée dans l'item, car les herbes des marais sont définies comme des « herbes qui absorbent les nutriments et les déchets de l'eau »

(cf. figure 3). Cette information utile pour répondre rend cet item dépendant aux ressources disponibles. Le contexte est à la fois sociétal et global car le problème de l'item présente des aspects sociétaux et dans une plus large mesure, environnementaux.

En termes de raisonnements et de stratégies de réponse, cet item présente, selon nous, un DOK de niveau 2¹⁶ bien que la réponse soit intégrée dans l'item, il ne suffit pas simplement de rappeler ou reconnaître une définition pour répondre. Il est d'abord nécessaire pour l'élève de se remémorer ses réponses à l'item précédent (où il s'agissait de placer chaque organisme dans le bon bassin) pour identifier

Figure 3. Exemple d'un item PISA Science 2015 libéré « ferme aquatique durable » (OCDE, 2016)

Ferme aquatique durable
Question 2 / 3
Utilisez les informations ci-dessous. Cliquez sur la réponse de votre choix.

Le schéma montre le plan d'une ferme aquatique expérimentale avec trois grands bassins. De l'eau salée filtrée est pompée dans l'océan avant de passer de bassin en bassin puis d'être finalement rejetée dans l'océan. L'objectif principal de cette ferme est d'élever des soles communes qui seront capturées selon des méthodes de pêche durable.

- **Sole commune** : le poisson de cet élevage. Sa nourriture préférée est la néréide.

Les organismes suivants sont également utilisés dans cette ferme :

- **Algues microscopiques** : organismes microscopiques qui ont besoin uniquement de lumière et de nutriments pour se développer.
- **Néréides** : invertébrés qui se développent très rapidement en se nourrissant d'algues microscopiques.
- **Coquillages** : organismes qui se nourrissent d'algues microscopiques et de petits organismes présents dans l'eau.
- **Herbes des marais** : herbes qui absorbent les nutriments et les déchets de l'eau.

Les chercheurs ont remarqué que l'eau rejetée dans l'océan contient une grande quantité de nutriments. Pour réduire ce problème, lequel des éléments suivants devra être ajouté en plus grande quantité dans la ferme aquatique ?

Nutriments
 Néréides
 Coquillages
 Herbes des marais

L'eau est rejetée dans l'océan

L'eau de l'océan entre dans les bassins

Des nutriments sont ajoutés dans ce bassin

Filtre

Filtre

Filtre

L'eau est nettoyée dans ce bassin

Les poissons sont capturés dans ce bassin

Filtres qui permettent uniquement aux algues microscopiques de traverser les bassins en suivant le courant d'eau

la relation entretenue par les différents organismes et leur rôle dans la ferme aquatique durable. Il faut ensuite comparer les organismes (proposés dans le choix de réponse) pour déterminer celui qui est le plus approprié pour résoudre le problème. Cet item nécessite donc un processus cognitif en plusieurs étapes avant de pouvoir y répondre.

La formulation du problème implique de la part de l'élève de se projeter vers une communauté de savoir c'est-à-dire vers une profession représentant une autorité scientifique ou plus exactement dans cet item en essayant d'adopter la posture de chercheur. On peut observer qu'un matching est possible et que celui-ci peut conduire l'élève vers une réponse erronée. En effet le terme « nutriments » est utilisé dans le libellé de l'item et correspond également au premier choix de réponse proposé. Le niveau de référence à la vie quotidienne est plutôt faible dans cet item en comparaison aux 183 autres items.

L'échantillon d'élèves français de 15 ans à avoir passé l'enquête en 2015 a obtenu un taux de réussite

de 72 %, ce qui correspond à un niveau de performance relativement élevé. Toutefois, si l'on s'intéresse maintenant aux résultats obtenus à cet item en fonction du SESC, on se rend compte que l'écart de performances entre les élèves les plus défavorisés (SESC1) et les plus favorisés (SESC4) est l'un des plus importants de l'échantillon des élèves français avec un écart de 27 points¹⁷.

On peut ainsi se rendre compte qu'il n'est pas facile de prime abord d'identifier quelles caractéristiques ont pu rendre cet item difficile pour certains élèves, plus que pour d'autres. Un modèle de difficulté de questions prend donc tout son sens pour tenter par exemple de comprendre quel élément ou combinaison d'éléments présents dans l'item a pu augmenter sa difficulté.

DISCUSSION – CONCLUSION

Cet article vise à présenter la façon dont nous avons élaboré notre modèle de difficultés des questions qui a pour but d'identifier des caractéristiques

Tableau 6. Caractéristiques du 2^e item de l'unité « ferme aquatique durable » (PISA Science 2015)

Catégorie	Caractéristique et sous-caractéristique	Type de caractéristique et sous-caractéristique Item 2 « Ferme aquatique durable »
Caractéristiques liées aux contenus en jeu dans la tâche	Type de connaissances	De contenu
	Systèmes <i>Croisement connaissances de contenu et système</i>	Systèmes vivants <i>Systèmes vivants des connaissances de contenu</i>
	Type de compétences	Interpréter des données et des faits de manière scientifique
	Champs d'application	Qualité de l'environnement
Caractéristiques intrinsèques de l'item – aspects formels	Format	Choix multiple simple
	Longueur du texte (nombre de mots)	238 (quartile 3= texte relativement long)
	Type d'illustration	Schéma
	Simulation	Absence de simulation
Caractéristiques intrinsèques de l'item – aspects liés au sens	Contextes	Sociétal/Global
	Réponse dans l'item	Présence de la réponse dans le texte (cf. figure 3)
	Dépendance ou Indépendance aux ressources	Dépendance aux ressources disponibles
	<i>Croisement dépendance aux ressources et contexte</i>	<i>Sociétal/Global</i>
	<i>Dépendance aux ressources et contexte selon sa forme (simple ou double)</i>	<i>Double</i>
	Lien question-illustration	Présence d'un lien question-illustration
	<i>Lien explicite ou implicite</i>	<i>Lien implicite</i>
Caractéristiques liées aux raisonnements et stratégies de réponse des élèves	Complexité cognitive (DOK)	Niveau 2
	Matching	Matching présent
	<i>Matching aidant ou invalidant</i>	<i>Matching invalidant</i>
	Projection	Présence d'une projection
	<i>Type de Projection</i>	<i>Projection vers une communauté de savoirs (c'est-à-dire une projection vers une profession représentant une autorité scientifique)</i>
	<i>Projection directe ou indirecte</i>	<i>Projection directe</i>
	Niveau de référence à la vie quotidienne	Référence à la vie quotidienne faible

de tâches de sciences influençant la variation des performances des élèves selon leur statut socio-économique et culturel et/ou leur niveau de performance. Ce travail nous a conduits d'une part, à une réflexion sur la construction de notre modèle et d'autre part, sur les attentes que ce modèle pourrait nous montrer comme résultats lorsqu'il sera testé statistiquement.

Construction du modèle

Lorsque nous avons établi notre modèle, pour lequel nous avons défini des caractéristiques des

items pouvant présenter des difficultés pour les élèves dont certaines possiblement spécifiques du milieu socio-économico-culturel et/ou du niveau de performance des élèves, nous avons fait appel à des types de travaux de différentes natures :

- des résultats empiriques provenant par exemple de grandes enquêtes internationales comme TIMSS ou PISA, donnant déjà des indications sur la difficulté de certaines modalités de caractéristiques, comme le format de l'item (les items avec un format réponse ouverte sont plus difficiles pour les élèves que les items avec un format choix multiple). Nous avons également utilisé cette approche

empirique pour la caractéristique de la présence de la réponse dans l'item ;

- du cadre théorique de l'évaluation (dans notre cas celui de PISA 2015) : Ce cadre explicite que l'évaluation porte sur les compétences à mettre en œuvre dans différents contextes, celles-ci étant influencées par les connaissances (scientifiques, procédurales et épistémiques). On retrouve ici les caractéristiques liées aux contenus en jeu dans la tâche ainsi qu'une des caractéristiques intrinsèques à l'item : les contextes ;
- des résultats de recherche dans le domaine des sciences du langage sur des questions de compréhension de l'écrit ou du lien texte-représentations ou encore sur la question des communautés discursives. Ces travaux nous ont conduits à préciser des caractéristiques aussi bien intrinsèques à la tâche (longueur du texte, type d'illustration, etc.) que liées aux raisonnements et stratégies de réponse (par exemple la projection) ;
- des résultats de recherches dans les domaines de la cognition, de la psychologie de l'éducation et/ou de la didactique des sciences, menées principalement à partir d'études de cas nous ont amenés à construire des caractéristiques liées aux raisonnements et stratégies (échelle de complexité cognitive, « matching », etc.) ;
- des résultats de recherches dans les domaines de la sociologie de l'éducation et en didactique des sciences et des mathématiques, portant sur la prégnance de l'expérience quotidienne pour les élèves en particulier ceux ayant un SESC faible et/ou peu performants.

Cette variété de nature et de domaines de recherche met en évidence la multiplicité de l'origine des difficultés possibles des élèves lorsqu'ils répondent à des questions d'évaluation.

Ce choix de caractéristiques que nous avons opéré, à partir de ces différents types de travaux pour la construction de ce modèle nous a fait réfléchir aux résultats auxquels nous pourrions nous attendre à partir de notre analyse statistique et des résultats PISA Science 2015.

Les résultats envisagés

Dans nos travaux antérieurs, nous avons observé et analysé les comportements verbaux et gestuels de

binômes d'élèves de niveaux de performance scolaire différents lorsqu'ils résolvent des items PISA Science 2006 (Le Hebel *et al.*, 2012 ; 2014 ; 2016). Ces études nous ont permis d'inférer les stratégies de réponse des élèves et nous avons constaté que les stratégies de réponse utilisées sont différentes selon leur niveau de performance scolaire.

Les élèves performants montrent une variété de capacités qui leur permet de diversifier leurs analyses et leurs stratégies de réponses pour discuter et répondre aux questions aboutissant en général à une réponse juste. À l'inverse, les élèves en difficulté construisent des représentations mentales instables de la situation proposée dans l'item et du but de la question, et un rapport au savoir faible, ils sont souvent limités à des stratégies peu ancrées sur le savoir comme la « recherche de mots similaires dans le texte introductif, la question ou les réponses proposées dans le cas d'un format de question à choix multiple » (matching), ou la « paraphrase dans le cas d'un format de questions à réponse ouverte », voire le « hasard ». Cela nous conduit à faire l'hypothèse que ce type de résultats devrait se traduire dans l'analyse statistique de notre modèle de difficulté par des caractéristiques discriminantes incluses par exemple dans la catégorie « contenus en jeu dans la tâche » impliquant un rapport au savoir différent des élèves selon leur niveau de performance.

Une autre hypothèse que nous faisons, en nous basant sur des recherches antérieures, est que les caractéristiques contenues dans la catégorie « stratégie et raisonnement » incluant notamment les caractéristiques « référence au quotidien » et « projection », probablement très dépendantes du contexte socio-économico-culturel des élèves, seront plus discriminantes selon le niveau SESC des élèves.

Ainsi, nous nous attendons à une conjonction de caractéristiques explicatives de la difficulté d'une tâche de science, qui selon notre hypothèse sera différente en termes de nature et de pondération de ces caractéristiques, pour chacune des quatre populations étudiées (bas SESC – faible performance, bas SESC – forte performance, haut SESC – faible performance, haut SESC – forte performance). Ceci serait notamment lié au fait que certaines caractéristiques soient plus fortement en lien avec le contexte socio-économico-culturel des élèves.

Implication pour ce type d'études pour l'enseignement et la conception d'évaluations en sciences

Les résultats de ce type d'études pourraient permettre de mieux comprendre les difficultés rencontrées par les élèves lors de la résolution d'une tâche de science selon leur performance scolaire et leur niveau socio-économique et culturel d'appartenance et ce, au-delà de tâches scientifiques telles que celles proposées par PISA. Ainsi, cela pourrait contribuer à aider les enseignants à mieux cibler ces difficultés dans leur pratique de classe et notamment à en tenir compte dans leurs propres conceptions d'évaluation.

On peut imaginer que les résultats de ce type d'études contribuent à la prise de conscience des enseignants vis-à-vis de certaines caractéristiques très défavorables pour des élèves de bas SESC et/ou de faible performance scolaire. Ces caractéristiques ne peuvent être visibles qu'à partir de travaux de recherche ciblés sur cette problématique, ce qui nécessite leur prise en compte spécifique dans la formation des enseignants. Par exemple, la référence à la vie quotidienne dans une tâche de sciences peut être pénalisante pour certains élèves, et sa prise en compte explicite pourrait aider ces élèves à faire la transition entre le monde de la vie quotidienne d'un élève et la science de l'école (Phelan *et al.*, 1991). Ceci renvoie à la nécessité d'une conceptualisation d'une frontière culturelle entre plusieurs micro-cultures (Aikenhead, 1996). De plus, si l'analyse statistique confirme le rôle discriminatoire de ces caractéristiques selon la performance et le SESC, une attention particulière devra être portée pour mettre à disposition les résultats. En effet la variété des disciplines de référence qui ont été mises en œuvre pour construire ces caractéristiques (didactique des sciences, psychologie, sociologie, sciences du langage, etc.) et leur possible enchevêtrement conduisent à la nécessité de construire des documents adaptés pour les enseignants pour rendre ces résultats opérationnels dans leur pratique professionnelle.

NOTES

1. Considérant le petit nombre d'items PISA libérés, notre choix pour en sélectionner un était assez limité.
2. Pour la France – Connaissances de contenu : 489 points; connaissances procédurales et épistémiques : 499 points.

3. Pour la France – Systèmes Physiques : 492 points, Systèmes Vivants : 496 points; Systèmes de la Terre et de l'Univers : 496 points.

4. Connaissances de contenu : 97 items; Connaissances procédurales : 60 items; Connaissances épistémiques : 26 items.

5. Notons que chacune de ces trois compétences majeures comprend cinq sous-compétences : 1. Évaluer et concevoir des recherches scientifiques (Identifier la question étudiée dans des recherches scientifiques données / Identifier des questions qui se prêtent à des recherches scientifiques / Proposer une façon d'étudier une question de manière scientifique / Évaluer les moyens d'étudier une question de manière scientifique / Décrire et évaluer la façon dont les scientifiques garantissent que leurs données sont fiables et que leurs explications sont objectives et généralisables); 2. Expliquer des phénomènes de manière scientifique (Se remémorer les connaissances scientifiques appropriées et les appliquer / Identifier, utiliser et générer des représentations et des modèles explicatifs / Faire et justifier des prévisions appropriées / Proposer des hypothèses explicatives / Expliquer les implications potentielles de connaissances scientifiques pour la société); 3. Interpréter des données et des faits de manière scientifique (Transposer des données d'une représentation dans une autre / Analyser et interpréter des données, et en tirer des conclusions appropriées / Identifier des hypothèses, des faits et des raisonnements dans des textes scientifiques / Faire la distinction entre des arguments basés sur des théories et des faits scientifiques et ceux basés sur d'autres considérations / Évaluer des faits et des arguments scientifiques de sources différentes [quotidiens, Internet, revues, etc.]).

6. Pour la France – Expliquer des phénomènes de manière scientifique : 488 points; Évaluer et concevoir des investigations scientifiques : 498 points; Interpréter des données et des faits de manière scientifique : 501 points.

7. Taux de réussite moyen aux items PISA Science 2015 selon le champ d'application : Risques : 0,5509 / Qualité de l'environnement : 0,5186 / Ressources naturelles : 0,4884 / Frontières Sciences et techniques : 0,4781 / Santé et maladies : 0,4779.

8. Taux de réussite moyen aux items PISA Science 2015 en fonction de leur format : Choix multiple simple : 0,5874 / Choix multiple complexe : 0,5047 / Réponse ouverte : 0,4104.

9. Par exemple, dans notre répartition de nombre de mots, le premier quartile correspond au nombre de mots en dessous duquel se situent 25 % des items.

10. Taux de réussite moyen aux items PISA Science 2015 pour les items faisant appel à la simulation : 0,4867 / Taux

de réussite moyen aux items PISA Science 2015 pour les items sans simulation : 0,4972.

11. Classification proposée par PISA Science 2015 pour les contextes : contextes personnels ; contextes locaux/nationaux ; contextes mondiaux.

12. Taux de réussite moyen aux items PISA Science 2015 pour les items relevant d'un contexte : global : 0,5408 / sociétal : 0,5197 / global-personnel : 0,4934 / sociétal-global : 0,4906 / sociétal-personnel : 0,4763.

13. Rappelons que les items PISA Science sont organisés en unités. Chaque unité comprend un texte introductif plus ou moins long, pouvant contenir ou pas une ou plusieurs illustrations, suivi de plusieurs items (de deux à cinq) dans lesquels sont posées les questions.

14. https://www.oecd-ilibrary.org/education/resultats-du-pisa-2015-volume-i/items-de-l-enquete-pisa-2015_9789264267534-15-fr (Annexe C, p. 487).

15. https://www.oecd-ilibrary.org/education/resultats-du-pisa-2015-volume-i/items-de-l-enquete-pisa-2015_9789264267534-15-fr (Annexe C, p. 485).

16. L'item 2 de « ferme aquatique durable » a été classé en DOK de niveau faible par PISA Science 2015 sur une échelle à trois niveaux (faible-moyen-fort). Nous avons recodé cet item à un niveau 2 de DOK sur l'échelle à quatre niveaux de Webb (2007).

17. Nous nous sommes concentrés sur les groupes extrêmes de l'échantillon afin d'observer le plus de différences possibles. En effet, les deux groupes contigus contrastent moins et les écarts sont ainsi potentiellement moins significatifs.

RÉFÉRENCES

- Ahmed, A. et Pollitt, A. (1999). *Curriculum Demands and Question Difficulty*. Paper presented at IAEA Conference, Slovenia, May.
- Aikenhead, G. S. (1996). Science Education: Border Crossing into the Subculture of Science. *Studies in Science Education*, 27(1), 1-52. <https://doi.org/10.1080/03057269608560077>.
- Ainsworth, S. (2008). How do animations influence learning? Dans D. H. Robinson et G. Schraw (eds.), *Recent innovations in educational technology that facilitate student learning*. Information Age Pub.
- Airey, J. et Linder, C. (2009). A disciplinary discourse perspective on university science learning: Achieving fluency in a critical constellation of modes. *Journal of Research in Science Teaching*, 46(1), 27-49. <https://doi.org/10.1002/tea.20265>.
- Bautier, É. et Rayou, P. (2013). La littératie scolaire : Exigences et malentendus. Les registres de travail des élèves. *Éducation & Didactique*, 7(7-2), 29-46. <https://doi.org/10.4000/educationdidactique.1721>.
- Bautier, É. et Rochex, J.-Y. (1997). Ces malentendus qui font les différences. Dans J.-P. Terrail (dir.), *La scolarisation de la France : Critique de l'état des lieux* (p. 105-122). La Dispute.
- Bergqvist, E. et Österholm, M. (2010). A theoretical model of the connection between the process of reading and the process of solving mathematical tasks. *Mathematics and Mathematics Education: Cultural and Social Dimensions. Proceeding of MADIF 7*, 47-57. <http://www.diva-portal.org/smash/record.jsf?pid=diva2:298097>.
- Bergqvist, E., Theens, F. et Österholm, M. (2018). The role of linguistic features when reading and solving mathematics tasks in different languages. *The Journal of Mathematical Behavior*, 51, 41-55. <https://doi.org/10.1016/j.jmathb.2018.06.009>.
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain* (2nd edition). Addison-Wesley Longman Ltd.
- Boaler, J. (1993). Encouraging the transfer of “school” mathematics to the “real world” through the integration of process and content, context and culture. *Educational Studies in Mathematics*, 25(4), 341-373. <https://doi.org/10.1007/BF01273906>.
- Bonnery, S. (2007). *Comprendre l'échec scolaire. Élèves en difficultés et dispositifs pédagogiques*. La Dispute.
- Carney, R. N. et Levin, J. R. (2002). Pictorial Illustrations Still Improve Students' Learning from Text. *Educational Psychology Review*, 14(1), 5-26. <https://doi.org/10.1023/A:1013176309260>.
- Cerdán, R., Gilabert, R. et Vidal-Abarca, E. (2011). Selecting information to answer questions: Strategic individual differences when searching texts. *Learning and Individual Differences*, 21(2), 201-205. <https://doi.org/10.1016/j.lindif.2010.11.007>.
- Chopin, H., Dubois, V., Galisson, A., Rouet, J.-F., Everard, J.-M. et Paquelin, D. (2001). Des nouveaux outils au processus d'innovation pédagogique : Qui est l'élève ? *Spirale – Revue de recherches en éducation*, 28(1), 123-144. <https://doi.org/10.3406/spira.2001.1458>.
- Crisp, V. et Grayson, R. (2013). Modelling question difficulty in an A level physics examination. *Research Papers in Education*, 28(3), 346-372. <https://doi.org/10.1080/02671522.2012.673005>.
- Dempster, E. R. (2007). Textual strategies for answering multiple choice questions among South African learners: What can we learn from TIMSS 2003? *African Journal of Research in Mathematics, Science and Technology Education*, 11(1), 47-60. <https://doi.org/10.1080/10288457.2007.10740611>.
- Dempster, E. R. (2006, 26 au 30 juin). *Strategies for answering multiple choice questions among south African learners: What can we learn from TIMSS 2003*. Proceedings for the 4th Sub-Regional Conference on Assessment in Education. Pretoria: UMALUSI.
- Dempster, E. R. et Reddy, V. (2007). Item readability and science achievement in TIMSS 2003 in South

- Africa. *Science Education*, 91(6), 906-925. <https://doi.org/10.1002/sce.20225>.
- Dhillon, D. (2011). *Predictive models of question difficulty – A critical review of the literature*. Manchester: AQA Centre for Education Research and Policy.
- Dong, D. (2020). What do We Know about Context: An Integrated Analysis of Context Characteristics of Science Assessment Items [Thesis]. <https://digital.lib.washington.edu/443/researchworks/handle/1773/45482>.
- Drouin, A.-M. (1987). Des images et des sciences. *Aster*, 4, 1-32. <https://doi.org/10.4267/2042/9175>.
- Dyrvold, A. (2020). Relations between semiotic resources in mathematics tasks: A source of students' difficulties. *Research in Mathematics Education*, 22(3), 265-283. <https://doi.org/10.1080/14794802.2019.1689160>.
- El Masri, Y. H., Ferrara, S., Foltz, P. W. et Baird, J.-A. (2017). Predicting item difficulty of science national curriculum tests: The case of key stage 2 assessments. *The Curriculum Journal*, 28(1), 59-82. <https://doi.org/10.1080/09585176.2016.1232201>.
- Fakcharoenphol, W., Morphey, J. W. et Mestre, J. P. (2015). Judgments of physics problem difficulty among experts and novices. *Physical Review Special Topics – Physics Education Research*, 11(2), 020128. <https://doi.org/10.1103/PhysRevSTPER.11.020128>.
- Felouzis, G. et Charmillot, S. (2012). Comment sont construites les enquêtes PISA? Dans G. Felouzis et S. Charmillot, *Les enquêtes PISA, Que sais-je?* (p. 13-46). Presses universitaires de France. https://www.cairn.info/feuilleter.php?ID_ARTICLE=PUF_FELOU_2012_01_0013.
- Ford, M. J. et Wargo, B. M. (2012). Dialogic framing of scientific content for conceptual and epistemic understanding. *Science Education*, 96(3), 369-391. <https://doi.org/10.1002/sce.20482>.
- Grapin, N. et Sayac, N. (2017). Évaluer la maîtrise de la numération écrite chiffrée : Choix du format QCM et validité d'items d'évaluations externes. *Éducation & Didactique*, 11(3), 55-72. <https://doi.org/10.4000/educationdidactique.2836>.
- Grisay, A. (2003). Translation procedures in OECD/PISA 2000 international assessment. *Language Testing*, 20(2), 225-240. <https://doi.org/10.1191/0265532203l12540a>.
- Hannus, M. et Hyönä, J. (1999). Utilization of Illustrations during Learning of Science Textbook Passages among Low- and High-Ability Children. *Contemporary Educational Psychology*, 24(2), 95-123. <https://doi.org/10.1006/ceps.1998.0987>.
- IEA. (2016). *TIMSS & PIRLS – TIMSS 2015* (Trends in International Mathematics and Science Study). <https://timssandpirls.bc.edu/timss2015/index.html>.
- Katz, I. R., Lipps, A. W. et Trafton, J. G. (2002). Factors Affecting Difficulty in the Generating Examples Item Type. *ETS Research Report Series*, 2002(1), i-40. <https://doi.org/10.1002/j.2333-8504.2002.tb01874.x>.
- Keehner, M., Gorin, J. S., Feng, G. et Katz, I. R. (2016). Developing and Validating Cognitive Models in Assessment. Dans A. A. Rupp, J. P. Leighton (eds.), *The Wiley Handbook of Cognition and Assessment* (p. 75-101). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118956588.ch4>.
- Lahire, B. (1995). *Tableaux de familles : Heurs et malheurs scolaires en milieux populaires*. Seuil Gallimard.
- Le Hebel, F., Montpied, P. et Tiberghien, A. (2012). Analyse de réponses d'élèves français à quelques questions de l'évaluation PISA sur la culture scientifique. *Recherches en didactiques*, 14(2), 65-83. <https://www.cairn.info/revue-recherches-en-didactiques-2012-2-page-65.htm>.
- Le Hebel, F., Montpied, P. et Tiberghien, A. (2014). Which Effective Competencies Do Students Use in PISA Assessment of Scientific Literacy? Dans C. Bruguière, A. Tiberghien et P. Clément (Éds.), *Topics and Trends in Current Science Education: 9th ESERA Conference Selected Contributions* (Springer, Dordrecht, p. 273-289). https://doi.org/10.1007/978-94-007-7281-6_17.
- Le Hebel, F., Montpied, P. et Tiberghien, A. (2016). Which Answering Strategies Do Low Achievers Use to Solve PISA Science Items? Dans *Insights from Research in Science Teaching and Learning* (p. 237-252). Springer, Cham. https://doi.org/10.1007/978-3-319-20074-3_16.
- Le Hebel, F., Montpied, P., Tiberghien, A. et Fontanieu, V. (2017). Sources of difficulty in assessment: Example of PISA Science items. *International Journal of Science Education*, 39(4), 468-487. <https://doi.org/10.1080/09500693.2017.1294784>.
- Le Hebel, F., Tiberghien, A., Montpied, P. et Fontanieu, V. (2019). Teacher prediction of student difficulties while solving a science inquiry task: Example of PISA Science items. *International Journal of Science Education*, 41(11), 1517-1540. <https://doi.org/10.1080/09500693.2019.1615150>.
- Lemke, J. L. (1990). *Talking Science: Language, Learning, and Values*. Praeger Publishers Inc.
- Lemke, J. L. (1998). *Teaching All the Languages of Science: Words, Symbols, Images, and Actions*. Conference on Science Education in Barcelona. <http://dx.doi.org/10.13140/2.1.4022.5608>.
- Liou, P.-Y. et Bulut, O. (2017). The Effects of Item Format and Cognitive Domain on Students' Science Performance in TIMSS 2011. *Research in Science Education*, 50(1), 99-121. <https://doi.org/10.1007/s11165-017-9682-7>.
- Marzano, R. J. et Kendall, J. S. (2007). *The New Taxonomy of Educational Objectives* (2nd Revised edition). SAGE Publications Inc.
- Maurice, J.-J. (1996). Une connaissance de l'élève dépendante des contraintes de l'action. *Revue française de pédagogie*, 114, 85-96. <https://www.jstor.org/stable/41200597>.
- McTigue, E. M. et Flowers, A. C. (2011). Science Visual Literacy: Learners' Perceptions and Knowledge of Diagrams. *The Reading Teacher*, 64(8), 578-589. <https://doi.org/10.1598/RT.64.8.3>

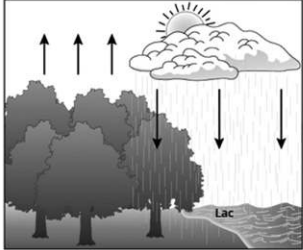
- Milenković, D., Segedinac, M., Hrin, T. et Gajić, G. (2015). Evaluation of Context-Level Effect on Students' Performance and Perceived Cognitive Load in Chemistry Problem-solving Tasks. *Croatian Journal of Education: Hrvatski Časopis Za Odgoj i Obrazovanje*, 17(4), 959-982. <https://hrcak.srce.hr/151223>.
- Mulholland, T. M., Pellegrino, J. W. et Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology*, 12(2), 252-284. [https://doi.org/10.1016/0010-0285\(80\)90011-0](https://doi.org/10.1016/0010-0285(80)90011-0).
- Mullis, I. V. S., Martin, M. O. et Foy, P. (2013). The Impact of Reading Ability on TIMSS Mathematics and Science Achievement at the Fourth Grade: An Analysis by Item Reading Demands. Dans M. O. Martin et I. V. S. Mullis (eds.), *TIMSS and PIRLS 2011: Relationships among Reading, Mathematics, and Science Achievement at the Fourth Grade – Implications For Early Learning* (p. 67-110). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- OCDE. (2016). *PISA 2015 Assessment and Analytical Framework – Science, Reading, Mathematic and Financial Literacy*. Éditions OCDE. <http://dx.doi.org/10.1787/9789264255425-en>.
- OCDE. (2016). *Résultats du PISA 2015 (Volume I) : L'excellence et l'équité dans l'éducation*. Éditions OCDE. https://www.oecd-ilibrary.org/fr/education/resultats-du-pisa-2015-volume-i_9789264267534-fr.
- Peeck, J. (1993). Increasing picture effects in learning from illustrated text. *Learning and Instruction*, 3(3), 227-238. [https://doi.org/10.1016/0959-4752\(93\)90006-L](https://doi.org/10.1016/0959-4752(93)90006-L).
- Pettersen, A. et Braeken, J. (2019). Mathematical Competency Demands of Assessment Items: A Search for Empirical Evidence. *International Journal of Science and Mathematics Education*, 17(2), 405-425. <https://doi.org/10.1007/s10763-017-9870-y>.
- Phelan, P., Davidson, A. L. et Cao, H. T. (1991). Students' Multiple Worlds: Negotiating the Boundaries of Family, Peer, and School Cultures. *Anthropology & Education Quarterly*, 22(3), 224-250. <https://doi.org/10.1525/aeq.1991.22.3.05x1051k>.
- Piel, S. et Schuchart, C. (2014). Social origin and success in answering mathematical word problems: The role of everyday knowledge. *International Journal of Educational Research*, 66, 22-34. <https://doi.org/10.1016/j.ijer.2014.02.003>.
- Pollitt, A. et Ahmed, A. (1999). *A New Model of the Question Answering Process*. Paper presented to the International Association for Educational Assessment, Bled, Slovenia, May.
- Pollitt, A., Entwistle, N., Hutchinson, C. et De Luca, C. (1985). *What Makes Exam Questions Difficult?: An Analysis of « O » Grade Questions and Answers*. Scottish Academic Pr.
- Prenzel, M., Häußler, P., Rost, J. et Senkbeil, M. (2002). Der PISA-Naturwissenschaftstest: Lassen sich die Aufgabenschwierigkeiten vorhersagen? *Unterrichtswissenschaft*, 30(2), 120-135.
- Rauch, D. P. et Hartig, J. (2010). Multiple-choice versus open-ended response formats of reading test items: A two-dimensional IRT analysis. *Psychological Test and Assessment Modeling*, 52(4), 354-379.
- Rebière, M. (2013). S'intéresser au langage dans l'enseignement des mathématiques, pour quoi faire? Dans A. Bronner, C. Bulf et C. Castela (dir.), *Questions vives en didactique des mathématiques : Problèmes de la profession d'enseignant, rôle du langage*. 16^e École d'été de didactique des mathématiques, Carcassonne, Languedoc-Roussillon, du 21 au 28 août 2011.
- Reid, D. J. et Beveridge, M. (1986). Effects of Text Illustration on Children's Learning of a School Science Topic. *British Journal of Educational Psychology*, 56(3), 294-303. <https://doi.org/10.1111/j.2044-8279.1986.tb03042.x>.
- Rémond, M. (2006). Éclairages des évaluations internationales PIRLS et PISA sur les élèves français. *Revue française de pédagogie. Recherches en éducation*, 157, 71-84. <https://doi.org/10.4000/rfp.433>.
- Rey-Mermet, A., Gade, M., Souza, A. S., von Bastian, C. C. et Oberauer, K. (2019). Is executive control related to working memory capacity and fluid intelligence? *Journal of Experimental Psychology: General*, 148(8), 1335-1372. <https://doi.org/10.1037/xge0000593>.
- Robinson, D. H. et Schraw, G. (eds.). (2008). *Recent innovations in educational technology that facilitate student learning*. Information Age Pub.
- Roth, W.-M. et Roychoudhury, A. (1993). The development of science process skills in authentic contexts. *Journal of Research in Science Teaching*, 30(2), 127-152. <https://doi.org/10.1002/tea.3660300203>.
- Schmitt, N., Jiang, X. et Grabe, W. (2011). The Percentage of Words Known in a Text and Reading Comprehension. *The Modern Language Journal*, 95(1), 26-43. <https://doi.org/10.1111/j.1540-4781.2011.01146.x>.
- Salgado, F. A. et Stacey, K. (2014). *Item Context Factors Affecting Students' Performance on Mathematics Items*. Mathematics Education Research Group of Australasia. Annual Meeting of the Mathematics Education Research Group of Australasia (MERGA), 37th, Sydney, New South Wales, Australia.
- Serder, M. et Jakobsson, A. (2016). Language Games and Meaning as Used in Student Encounters With Scientific Literacy Test Items. *Science Education*, 100(2), 321-343. <https://doi.org/10.1002/sce.21199>.
- Simkin, M. G. et Kuechler, W. L. (2005). Multiple-Choice Tests and Student Understanding: What Is the Connection? *Decision Sciences Journal of Innovative Education*, 3(1), 73-98. <https://doi.org/10.1111/j.1540-4609.2005.00053.x>.
- Slough, S. W., McTigue, E. M., Kim, S. et Jennings, S. K. (2010). Science Textbooks' Use of Graphical Representation: A Descriptive Analysis of Four Sixth Grade Science Texts. *Reading Psychology*, 31(3), 301-325. <https://doi.org/10.1080/02702710903256502>.

- Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities* (p. XI, 348). Lawrence Erlbaum.
- Stiller, J., Hartmann, S., Mathesius, S., Straube, P., Tiemann, R., Nordmeier, V., Krüger, D. et Upmeyer zu Belzen, A. (2016). Assessing scientific reasoning: A comprehensive evaluation of item features that affect item difficulty. *Assessment & Evaluation in Higher Education*, 41(5), 721-732. <https://doi.org/10.1080/02602938.2016.1164830>.
- Südkamp, A., Kaiser, J. et Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743-762. <https://doi.org/10.1037/a0027627>.
- Tout, D., Gal, I., Van Groenestijn, M., Manly, M. et Schmitt, M. J. (2020). PIAAC Numeracy Task Complexity Schema: Factors that impact on item difficulty. *Transition and Post-School Education and Training*. <https://10.37517/978-1-74286-609-3>.
- Turner, R., Dossey, J., Blum, W. et Niss, M. (2013). Using Mathematical Competencies to Predict Item Difficulty in PISA: A MEG Study. Dans M. Prenzel, M. Kobarg, K. Schöps, et S. Rönnebeck (dir.), *Research on PISA: Research Outcomes of the PISA Research Conference 2009* (p. 23-37). Springer Netherlands.
- Van Dijk, T. A. et Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.
- Vianin, P. (2009). Chapitre 1. Introduction. Dans *L'aide stratégique aux élèves en difficulté scolaire* (p. 21-45). De Boeck Supérieur.
- Vicente, S. et Machado, E. (2016). Arithmetic word problem solving. Are authentic word problems easier to solve than standard ones? *Journal for the Study of Education and Development*, 39(2), 349-379. <https://doi.org/10.1080/02103702.2016.1138717>.
- Vidal-Abarca, E. et Sanjose, V. (1998). Levels of comprehension of scientific prose: The role of text variables. *Learning and Instruction*, 8(3), 215-233. [https://doi.org/10.1016/S0959-4752\(97\)00020-0](https://doi.org/10.1016/S0959-4752(97)00020-0).
- Webb, N. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, 20(1), 7-25. https://doi.org/10.1207/s15324818ame2001_2.
- White, S. (2012). Mining the Text: 34 Text Features That Can Ease or Obstruct Text Comprehension and Use. *Literacy Research and Instruction*, 51(2), 143-164. <https://doi.org/10.1080/19388071.2011.553023>.
- Wijaya, A., Van den Heuvel-Panhuizen, M. et Doorman, M. (2015). Opportunity-to-learn context-based tasks provided by mathematics textbooks. *Educational Studies in Mathematics*, 89(1), 41-65.
- Wijaya, A., Van den Heuvel-Panhuizen, M., Doorman, M. et Robitzsch, A. (2014). Difficulties in solving context-based PISA mathematics tasks: An analysis of students' errors. *The Mathematics Enthusiast*, 11, 555-584.
- Yildirim, H. H. et Ceylan, E. (2020). SES Related Differential Factors in Turkey's PISA 2015 Science Literacy Data. *Ankara University Journal of Educational Sciences*, 53(3), 931-951.

ANNEXE 1

Item illustrant le « matching invalidant »

1 L'illustration ci-dessous montre le cycle de l'eau dans l'écosystème d'une forêt.



Les flèches qui pointent vers le bas montrent la pluie qui tombe dans la forêt. Une certaine quantité d'eau de pluie est absorbée par les arbres à partir du sol.

Quel processus du cycle de l'eau est présenté par les flèches qui pointent vers le haut ?

- A l'absorption de l'eau par les racines des arbres
- B la production de dioxyde de carbone par les arbres
- C l'évaporation de l'eau des feuilles des arbres dans l'air
- D le rejet de dioxyde de carbone dans l'air par les animaux vivant dans les arbres

	Réponse A	Réponse B	Réponse C	Réponse D
% d'élèves	25,5	16,1	53,6	3,5

Source : TIMSS-grade 8 – classe de 4^e – Exemples d'items de sciences et vie de la Terre

ANNEXE 2

Item illustrant la « projection directe vers une communauté de savoir »

CULTURES GÉNÉTIQUEMENT MODIFIÉES

LE MAÏS OGM DEVRAIT ÊTRE INTERDIT

Des groupes de protection de la nature ont demandé l'interdiction d'une nouvelle espèce de maïs génétiquement modifiée (OGM, organisme génétiquement modifié).

Ce maïs OGM est conçu pour résister à un nouvel herbicide puissant qui détruit les plants de maïs traditionnels. Ce nouvel herbicide détruira la plupart des mauvaises herbes qui poussent dans les champs de maïs.

Les protecteurs de la nature déclarent que, comme ces mauvaises herbes sont une source de nourriture pour les petits animaux, en particulier les insectes, l'utilisation de ce nouvel herbicide avec le maïs OGM nuira à l'environnement. Les partisans du maïs OGM répondent qu'une étude scientifique a démontré que cela n'arrivera pas.

Source : items libérés PISA Science 2006

Voici quelques détails de l'étude scientifique mentionnée dans l'article ci-dessus :

On a semé du maïs dans 200 champs à travers le pays.

On a divisé chaque champ en deux parties. Dans une moitié, on a cultivé du maïs génétiquement modifié (OGM) traité avec le nouvel herbicide puissant, et dans l'autre moitié on a cultivé du maïs traditionnel traité avec un herbicide traditionnel.

On a trouvé à peu près le même nombre d'insectes sur le maïs OGM traité avec le nouvel herbicide que sur le maïs traditionnel traité avec l'herbicide traditionnel.

Question 1 : CULTURES GÉNÉTIQUEMENT MODIFIÉES

Dans l'étude scientifique mentionnée par l'article, quels sont les facteurs qu'on a volontairement fait varier ? Entourez « Oui » ou « Non » pour chacun des facteurs suivants.

Est-ce que, dans l'étude, on a volontairement fait varier ce facteur ?	Oui ou Non ?
Le nombre d'insectes dans l'environnement	Oui / Non
Les types d'herbicide utilisés	Oui / Non

Crédit complet : Les deux réponses sont correctes. Dans l'ordre : Non, Oui.

Question 2 : CULTURES GÉNÉTIQUEMENT MODIFIÉES

On a semé du maïs dans 200 champs à travers le pays. Pourquoi les scientifiques ont-ils utilisé plus d'un site ?
A Afin que de nombreux agriculteurs puissent essayer le nouveau maïs OGM.

B Pour voir quelle quantité de maïs OGM ils pourraient cultiver.

C Pour recouvrir le plus de terrain possible avec des cultures OGM.

D Pour inclure diverses conditions de culture du maïs.

Crédit complet : D. Pour inclure diverses conditions de culture du maïs.

ANNEXE 3

Item illustrant la projection vers une communauté de professions représentant une autorité scientifique

7 Comprendre le rôle de la pectine chez les plantes

✓ Interpréter des résultats et en tirer des conclusions

Les cellules végétales sont entourées d'une paroi constituée d'un réseau complexe de molécules : on y trouve, entre autres, de la pectine. Des chercheurs ont identifié deux mutants, QUASIMODO (qua1 et qua2), qui ont moins de pectine dans leur paroi que les plantes non mutées (dites « sauvages »).



1 Épiderme de la tige chez une plante sauvage et une plante QUASIMODO âgées de deux semaines (MET).



2 Taille de la tige après 4 jours de germination.

Questions

- À partir des résultats d'expériences, proposez une explication à la taille des mutants QUASIMODO.
- Expliquez en quoi ces résultats permettent de mieux comprendre le rôle de la pectine dans la paroi végétale.