



HAL
open science

La motivation des élèves français face à des évaluations à faibles enjeux

Saskia Keskpaik, Thierry Rocher

► To cite this version:

Saskia Keskpaik, Thierry Rocher. La motivation des élèves français face à des évaluations à faibles enjeux : comment la mesurer ? son impact sur les réponses. *Éducation & formations*, 2015, Évaluation des acquis : principes, méthodologie, résultats, n° 86-87 (06), pp. 119-139. 10.48464/ef-86-87-06 . halshs-03572939

HAL Id: halshs-03572939

<https://shs.hal.science/halshs-03572939>

Submitted on 14 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

LA MOTIVATION DES ÉLÈVES FRANÇAIS FACE À DES ÉVALUATIONS À FAIBLES ENJEUX

Comment la mesurer ? Son impact sur les réponses¹

Saskia Keskpaik et Thierry Rocher
MENESR-DEPP, bureau de l'évaluation des élèves

Les évaluations standardisées des élèves, telles que Cedre ou PISA, renvoient à des enjeux politiques croissants, alors qu'elles restent à faibles enjeux pour les élèves participants. Dans le système éducatif français, où la notation tient une place prépondérante, la question de la motivation des élèves face à ces évaluations mérite d'être posée. En 2011, afin d'explorer cette question, une expérience a été menée en France à partir du test PISA. Suite à cette expérience, un instrument pour mesurer la motivation a été adapté à partir du « thermomètre d'effort » proposé dans PISA. Cet instrument a été introduit dans plusieurs évaluations conduites au niveau national par la DEPP, sur des échantillons de plusieurs milliers d'élèves, en fin de primaire (CM2) et en fin de collège (troisième). Ces données permettent de distinguer la motivation de l'élève de la difficulté perçue du test, et ainsi de mieux appréhender le lien entre la motivation des élèves français et leur performance. L'analyse de ces données renseigne en outre sur le rôle de certaines caractéristiques, des élèves ou des évaluations elles-mêmes, dans le degré de motivation à répondre aux questions de l'évaluation.

Dans le système éducatif français où la notation tient une place prépondérante, la question de la motivation des élèves face à une évaluation sans enjeux pour eux mérite d'être posée. En effet, si les élèves ne sont pas motivés à faire de leur mieux lors de telles évaluations, la validité des résultats et leur interprétation peuvent être interrogées. De même, lorsque certains élèves ou des sous-populations d'élèves s'avèrent systématiquement moins motivés que d'autres, la comparabilité des résultats risque d'être biaisée.

1. Cet article fait suite à une communication lors du congrès international « Actualité de la Recherche en Éducation et en Formation » (AREF) à Montpellier en août 2013.

Plusieurs travaux ont été mis en œuvre, ces dernières années, pour étudier la motivation des élèves à répondre à des tests à faibles enjeux pour eux ainsi que la relation entre cette motivation et la performance [BUTLER et ADAMS, 2007 ; EKLÖF, 2008 ; O'NEIL *et alii*, 2004 ; PENK, POEHLMANN, ROPPELT, 2013]. Ces études, divergentes dans leurs conclusions, varient considérablement selon les méthodes utilisées ainsi que selon les instruments employés pour mesurer la motivation. Malgré un intérêt croissant pour le sujet, peu de tentatives ont été menées afin de construire une mesure valide de la motivation à répondre à des tests [NDINGA et FRENETTE, 2010, EKLÖF, 2008].

En 2011, une expérience a été menée en France à partir du pré-test de PISA 2012² [KESKPAIK et ROCHER, 2012]. Suite à cette expérience, un instrument pour mesurer la motivation a été adapté du « thermomètre d'effort » de PISA. Cet instrument a ensuite été introduit dans de nombreuses évaluations conduites, en 2012, au niveau national par la DEPP³ sur des échantillons de plusieurs milliers d'élèves, en fin de primaire (CM2) et en fin de collège (troisième). La validité de cet instrument a été interrogée lors d'une étude qualitative menée par la DEPP en mai 2013 ▶ **Encadré p. 123**.

La présente étude s'intéresse aux résultats de ces évaluations, notamment aux réponses des élèves aux questions relatives à la motivation, et à la variation de ces réponses selon diverses caractéristiques des évaluations, des élèves et des établissements. Nous déterminons des profils d'élèves et étudions le lien entre la motivation des élèves face à ces évaluations et les scores qu'ils y ont obtenus.

MESURE DE MOTIVATION : « EFFORT » VERSUS « APPLICATION »

Effort, application et difficulté

PISA mesure l'investissement des élèves face au test à l'aide d'un « thermomètre d'effort » ▶ **Figure 1**. Une étude exploratoire des données de ce thermomètre [KESKPAIK et ROCHER, 2012] nous a suggéré que le terme « effort » est susceptible de poser un problème d'interprétation, en mélangeant la motivation de l'élève avec la difficulté du test. Ainsi, les élèves performants ont pu déclarer faire peu « d'effort », car les exercices proposés leur ont paru faciles. Suite à cette étude, un instrument de mesure de la motivation a été adapté à partir du « thermomètre » de PISA. Plus précisément, les énoncés des questions ont été simplifiés pour réduire la charge de lecture, le terme « effort » a été remplacé par le terme « application », les échelles ont été placées horizontalement, et une échelle supplémentaire a été ajoutée afin d'interroger les élèves sur la difficulté des exercices proposés. Cet instrument varie légèrement selon le niveau scolaire. Au collège, les échelles sont sur dix positions ▶ **Figure 2**. En revanche, à l'école, une échelle sur quatre positions a été préférée, jugée plus adaptée aux élèves de l'école primaire.

L'analyse des réponses d'élèves à ces items « d'application » indique tout d'abord que la motivation au test (question 2 de l'instrument) est liée à la difficulté perçue du test

2. Dans chaque cycle PISA un pré-test, appelé *Field Trial*, est organisé un an avant le test principal (*Main Study*) afin de tester les items ainsi que les procédures de passation.

3. Direction de l'évaluation, de la prospective et de la performance (DEPP), ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche.

► **Figure 1** Thermomètre d'effort de PISA

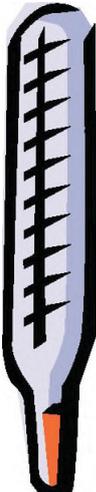
Quel effort avez-vous fourni pour répondre à ce test ?

Essayez de vous imaginer face à une situation de la vie réelle (à l'école ou dans un autre contexte) qui est très importante pour vous personnellement. Vous auriez envie de bien réussir, et pour cela, vous tentez de donner le meilleur de vous-même, en y consacrant le plus d'efforts possibles.

Dans cette situation, vous cochez la valeur la plus élevée sur le « thermomètre de l'effort », comme ci-dessous :

Par rapport à la situation que vous venez d'imaginer, quel effort pensez-vous avoir fourni en répondant à ce test ?

Si la note obtenue lors de ce test comptait pour votre bulletin scolaire, quel effort auriez-vous fourni ?



<input checked="" type="checkbox"/> 10	<input type="checkbox"/> 10	<input type="checkbox"/> 10
<input type="checkbox"/> 9	<input type="checkbox"/> 9	<input type="checkbox"/> 9
<input type="checkbox"/> 8	<input type="checkbox"/> 8	<input type="checkbox"/> 8
<input type="checkbox"/> 7	<input type="checkbox"/> 7	<input type="checkbox"/> 7
<input type="checkbox"/> 6	<input type="checkbox"/> 6	<input type="checkbox"/> 6
<input type="checkbox"/> 5	<input type="checkbox"/> 5	<input type="checkbox"/> 5
<input type="checkbox"/> 4	<input type="checkbox"/> 4	<input type="checkbox"/> 4
<input type="checkbox"/> 3	<input type="checkbox"/> 3	<input type="checkbox"/> 3
<input type="checkbox"/> 2	<input type="checkbox"/> 2	<input type="checkbox"/> 2
<input type="checkbox"/> 1	<input type="checkbox"/> 1	<input type="checkbox"/> 1

► **Figure 2** Instrument de mesure de la motivation au test (niveau collège)

Trois questions portant sur ce cahier

1. Sur une échelle de difficulté allant de 1 à 10, comment avez-vous trouvé les exercices de cette évaluation?

Très faciles Très difficiles

1 2 3 4 5 6 7 8 9 10

2. Comment vous êtes-vous appliqué(e) pour faire cette évaluation?

(indiquez votre degré d'application sur une échelle allant de 1 à 10)

Je ne me suis pas du tout appliqué(e) Je me suis énormément appliqué(e)

1 2 3 4 5 6 7 8 9 10

3. Si les résultats de cette évaluation comptaient pour votre bulletin scolaire, comment vous seriez-vous appliqué(e) ?

(indiquez votre degré d'application sur une échelle allant de 1 à 10)

Je ne me serais pas du tout appliqué(e) Je me serais énormément appliqué(e)

1 2 3 4 5 6 7 8 9 10

(question 1) et que ce lien est négatif. Plus les élèves ont jugé l'évaluation difficile, moins ils déclarent s'être appliqués pour la faire ▶ **Tableau 1**. Lorsque l'on étudie ce lien dans PISA, on note que le coefficient de corrélation est positif. Cela signifie que plus les exercices sont jugés difficiles par les élèves, plus ceux-ci déclarent avoir fourni d'effort pour y répondre. Les élèves qui disent avoir fourni beaucoup d'effort en répondant au test l'ont-ils fait car ils se sentaient très investis ou parce que la difficulté des exercices leur a demandé un effort considérable ? Le terme « effort » semble en effet ambigu pour informer sur la motivation des élèves face au test.

▶ **Tableau 1** Relation entre la difficulté perçue du test et la motivation au test

Évaluation	Coefficient de corrélation
Cedre histoire-géographie (3 ^e)	- 0,109
Socle compétences 1 et 3 (3 ^e)	- 0,053
Cedre histoire-géographie (CM2)	- 0,178
Cedre sciences (CM2)	- 0,181
Socle compétence 1 (CM2)	- 0,324
Socle compétence 2 (CM2)	- 0,281
Socle compétence 3 (CM2)	- 0,213
PISA (élèves de 15 ans)	0,069

Lecture : dans l'évaluation Cedre histoire-géographie niveau collège, la difficulté perçue du test est liée négativement à la motivation au test (coefficient de corrélation - 0,109) alors que dans PISA, où la motivation est mesurée en termes d'effort, ce lien est positif (0,069).

Note : les corrélations significatives au niveau 0,01 sont indiquées en gras.

Champs : France métropolitaine, public et privé sous contrat (Cedre) ; France métropolitaine + DOM, public et privé sous contrat (socle) ; France métropolitaine + DOM (sauf La Réunion), public et privé sous contrat (PISA).

Sources : MENESR-DEPP, évaluations Cedre et socle ; OCDE-PISA 2012.

La difficulté perçue des exercices entretient ainsi une relation avec la motivation à répondre au test, relation qui est par ailleurs plus forte au niveau primaire qu'au niveau secondaire. Ceci s'observe également au niveau agrégé : les évaluations jugées, en moyenne, les plus faciles, sont à la fois celles pour lesquelles le niveau de motivation des élèves s'avère le plus élevé. Dans la recherche des éléments d'explication, on peut noter le rôle du format d'exercices – les évaluations jugées les plus faciles et caractérisées par un degré plus grand de la motivation déclarée par les élèves sont celles qui se composent principalement des questions à choix multiples (QCM). Les informations qualitatives recueillies à l'aide des entretiens collectifs avec des élèves confirment cette observation : les élèves sont unanimes lorsqu'ils disent préférer les QCM aux questions à réponse construite ▶ **Encadré ci-contre**.

Profils de motivation

BUTLER et ADAMS [2007] proposent une analyse de l'investissement différentiel des élèves, et de l'effet de cet investissement sur la performance, en construisant un indicateur d'effort relatif à partir du « thermomètre d'effort » de PISA. Les auteurs

ÉTUDE QUALITATIVE DES CONDITIONS DE PASSATION

Une étude qualitative a été menée par la DEPP en mai 2013. Cette étude consistait à observer des conditions de passation et à conduire des entretiens collectifs (des *focus group*) avec des élèves dans une dizaine de collèges participant à une évaluation en sciences expérimentales [BOBINEAU, 2013]. L'objectif de l'étude était de connaître l'avis des élèves sur différents aspects de l'évaluation ainsi que d'obtenir des informations d'ordre qualitatif concernant leur investissement.

L'étude interrogeait aussi la validité de l'instrument de mesure de la motivation adapté à partir du « thermomètre d'effort ». Les renseignements provenant des entretiens collectifs avec des élèves montrent que ceux-ci ont bien compris les trois questions relatives au test qui leur étaient posées à la fin de l'évaluation

[figure 2 p.121]. La compréhension de l'instrument ne semble ainsi pas prêter à confusion. En outre, les élèves ont eu la possibilité de donner leur avis et de proposer des améliorations pour ces trois questions. De manière assez collégiale, ils proposent de réduire les échelles. Plus précisément, de les ramener à cinq possibilités de réponses, avec un milieu identifiable, deux extrêmes mais aussi la possibilité de mitiger leur réponse grâce aux entre-deux.

L'étude qualitative a mis en évidence le rôle que peut jouer le personnel de l'établissement dans la propension des élèves à participer à l'évaluation. Les entretiens conduits dans des établissements ayant pris le soin d'informer les élèves sur l'importance et l'utilité de l'évaluation ont révélé que les élèves de ces collèges comprenaient mieux pourquoi ils passaient ce test, sans conséquence directe pour eux.

classent les élèves en fonction de l'écart entre l'effort que ceux-ci déclarent avoir fourni en répondant au test PISA et l'effort qu'ils auraient fourni si les résultats du test avaient compté pour leur bulletin scolaire.

Cet indicateur très intéressant se prête aussi à quelques critiques. Notamment, il ne prend pas en compte le niveau général d'investissement des élèves. À titre d'exemple, un élève qui déclare avoir fourni un effort de 5 en répondant au test PISA et qui dit qu'il aurait fourni un effort de 7 si la note avait compté pour son bulletin a le même score sur l'échelle d'effort relatif qu'un élève qui coche respectivement 8 et 10. Ces deux élèves ont un écart de deux points entre les deux échelles et auront un score de l'effort relatif égal à 8 sur 10 (le score 10 représentant un même effort pour PISA que pour un test noté). Or, on peut supposer que pour un même effort relatif, deux élèves peuvent indiquer des niveaux de motivation générale différents, pouvant aboutir à des performances différentes au test.

Nous avons voulu vérifier cette hypothèse en positionnant les élèves simultanément sur les deux échelles – investissement au test et investissement si la note au test avait compté – et en étudiant leur score moyen au test en fonction de cette position. Observons les deux figures suivantes qui représentent les données des évaluations Cedre ⁴ histoire-géographie 2012 et PISA 2006 ► **Figures 3 et 4 p. 124-125**. Notons tout d'abord que la majorité des élèves se situent sur des niveaux élevés des échelles (7 points ou plus) et le plus souvent sous la ligne bissectrice, ce qui signifie qu'ils ont déclaré un investissement moins important pour un test standardisé que pour une épreuve notée. Le plus grand nombre a coché 10 sur l'échelle « si noté » et 8 sur l'échelle « test » (490

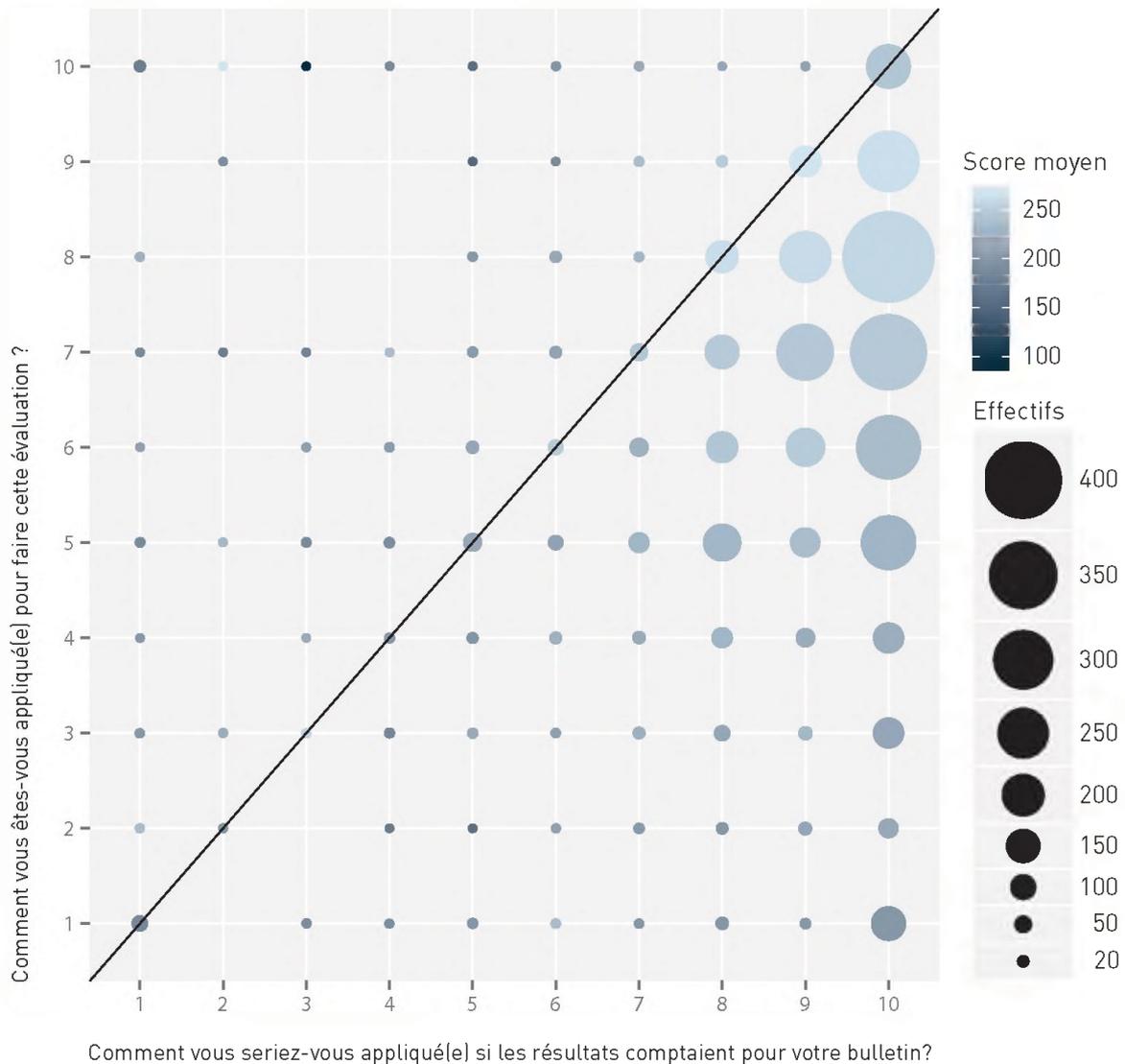
4. L'évaluation en histoire-géographie fait partie du cycle des évaluations disciplinaires réalisées sur échantillons (Cedre) que la DEPP a commencé à mettre en place en 2003 afin de rendre compte des résultats du système éducatif français au regard des objectifs fixés par les programmes [TROSSEILLE et ROCHER, ce numéro, p. 15].

élèves dans Cedre, soit 10 % des répondants, et 576 élèves dans PISA, soit 12 % des répondants).

On observe des variations de score considérables entre les élèves manifestant le même investissement relatif. Reprenons l'exemple évoqué plus haut et considérons deux groupes d'élèves qui ont coché respectivement sur les deux échelles 5 et 7 pour les uns, et 8 et 10 pour les autres. Si le score moyen en Cedre histoire-géographie est de 234 pour le premier groupe (n = 68), il s'élève à 260 pour le deuxième (n = 490), ce qui correspond à une différence d'environ un demi écart-type. Les données provenant de PISA montrent la même tendance : ces deux groupes d'élèves ont obtenu respectivement un score moyen de 490 (n = 46) et de 517 (n = 576) en 2006, soit une différence d'environ un tiers d'écart-type.

Nous nous concentrerons désormais sur l'enseignement secondaire où la motivation

► **Figure 3** Score moyen à l'évaluation Cedre histoire-géographie selon le niveau d'investissement (en termes « d'application »)



Lecture : 490 élèves ont coché 8 sur l'échelle « d'application » au test et 10 sur l'échelle « d'application » au test si les résultats avaient compté pour leur bulletin scolaire. Le score moyen en histoire-géographie de ce groupe d'élèves est de 260 points.

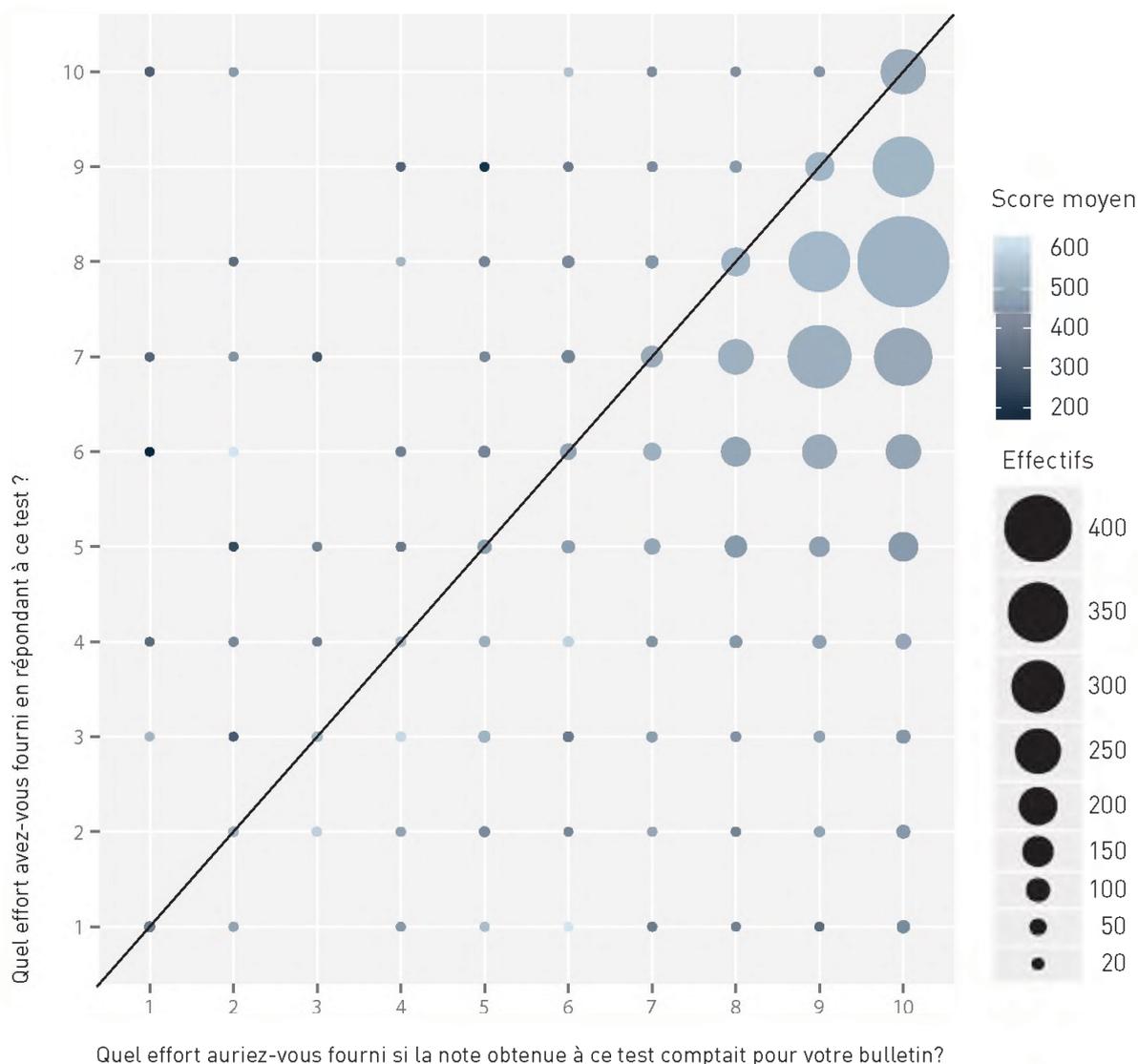
Champ : France métropolitaine, public et privé sous contrat.

Source : MENESR-DEPP, évaluation Cedre histoire-géographie 2012.

des élèves semble davantage poser problème qu'à l'école. En nous inspirant du travail de BUTLER et ADAMS [op. cit.], nous avons regroupé les élèves en fonction de leur position sur ces deux échelles. Des regroupements ont été testés en vue de réunir des élèves ayant des niveaux de motivation et des performances similaires mais aussi pour avoir des effectifs suffisamment importants pour pouvoir effectuer des analyses secondaires en croisant ces groupes avec d'autres variables. Le regroupement retenu vise à prendre en compte le niveau général des élèves et à assembler ceux qui se ressemblent à l'égard de l'interaction entre la motivation et le score.

Nous avons réparti les élèves en sept « profils de motivation » ► **Figure 5**. Comme BUTLER et ADAMS, nous nommons « irréalistes » les élèves qui déclarent s'être davantage appliqués pour faire l'évaluation standardisée qu'ils ne l'auraient fait si la note au test avait compté pour leur bulletin scolaire (donc tous ceux qui se trouvent au-dessus de la

► **Figure 4** Score moyen à l'évaluation PISA selon le niveau d'investissement (en termes « d'effort »)



Lecture : 576 élèves ont coché 8 sur l'échelle d'effort fourni au test et 10 sur l'échelle d'effort fourni si la note au test avait compté pour leur bulletin scolaire. Le score moyen en culture scientifique de ce groupe d'élèves est de 517 points.

Champ : France métropolitaine + DOM (sauf La Réunion), public et privé sous contrat.

Source : OCDE-PISA 2006.

diagonale exprimant un degré de motivation égal pour le test et pour une épreuve notée). Nous appelons « démotivés » ceux qui se considèrent aussi peu motivés pour le test standardisé que pour une épreuve notée (scores de 1 à 7 sur les deux échelles). Les « pragmatiques » sont les élèves qui se disent très investis dans une évaluation lorsque la note obtenue compte (scores de 8 à 10) mais très peu motivés pour un test à faibles enjeux (scores de 1 à 5). Les « peu motivés » sont ceux qui déclarent s'appliquer pour faire une évaluation qui compte (scores de 8 à 10) mais qui le font moins si les résultats n'ont pas de conséquence directe pour eux (scores 6 et 7). Enfin, les élèves motivés de manière générale sont appelés « réalistes » lorsque la différence entre les deux échelles est de 2 points, « assidus » si cet écart est égal à un point et « partisans » dans le cas où il n'y a pas de différence.

La présente étude se concentre sur la motivation des élèves face à des évaluations mais il serait très intéressant de comparer notre classification à la théorie de l'autodétermination et aux échelles de motivation développées antérieurement [DECI et RYAN, 1985]. Ainsi, on pourrait qualifier de « pragmatiques » les élèves qui sont motivés de manière extrinsèque, et notamment ceux en régulation externe. De même, les « démotivés » font fortement penser à l'amotivation. De telles comparaisons pourraient faire l'objet d'études ultérieures où les deux types d'échelles seraient proposés aux élèves.

La répartition de ces profils varie selon l'évaluation. Si les « pragmatiques » sont les plus représentés dans Cedre histoire-géographie (HG) ainsi que dans la session 5 d'EIST (en troisième), et les « peu motivés » en session 4 d'EIST (quatrième)⁵, ce sont les « assidus » qui prévalent dans l'évaluation du socle

► **Tableau 2** Répartition des profils de motivation selon l'évaluation (en %)

	Cedre HG	Socle	PISA	EIST session 4	EIST session 5
Irréalistes	5	4	5	6	5
Démotivés	10	9	10	9	8
Pragmatiques	26	16	12	20	27
Peu motivés	20	14	19	23	25
Réalistes	15	18	22	19	16
Assidus	14	23	21	14	12
Partisans	10	15	12	9	7
Non-réponse	12	4	8	21	12

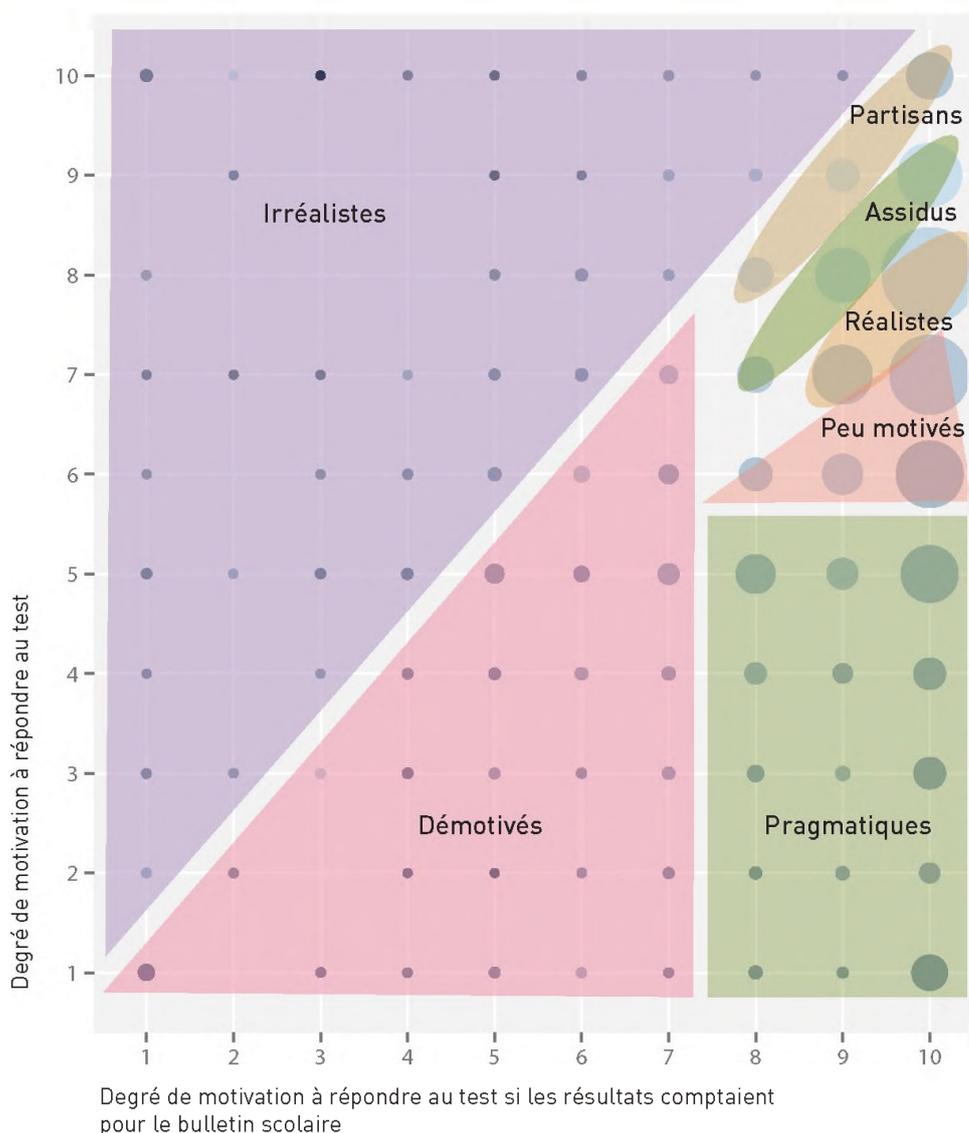
Lecture : dans l'évaluation Cedre histoire-géographie (HG) la part des élèves « démotivés » est de 10 %.

Note : la somme est égale à 100 % des répondants.

Champs : France métropolitaine, public et privé sous contrat (Cedre) ; France métropolitaine + DOM, public et privé sous contrat (Socle) ; France métropolitaine + DOM (sauf La Réunion), public et privé sous contrat (PISA) ; France métropolitaine, public (EIST).

Sources : MENESR-DEPP, évaluations Cedre, socle et EIST ; OCDE-PISA 2006.

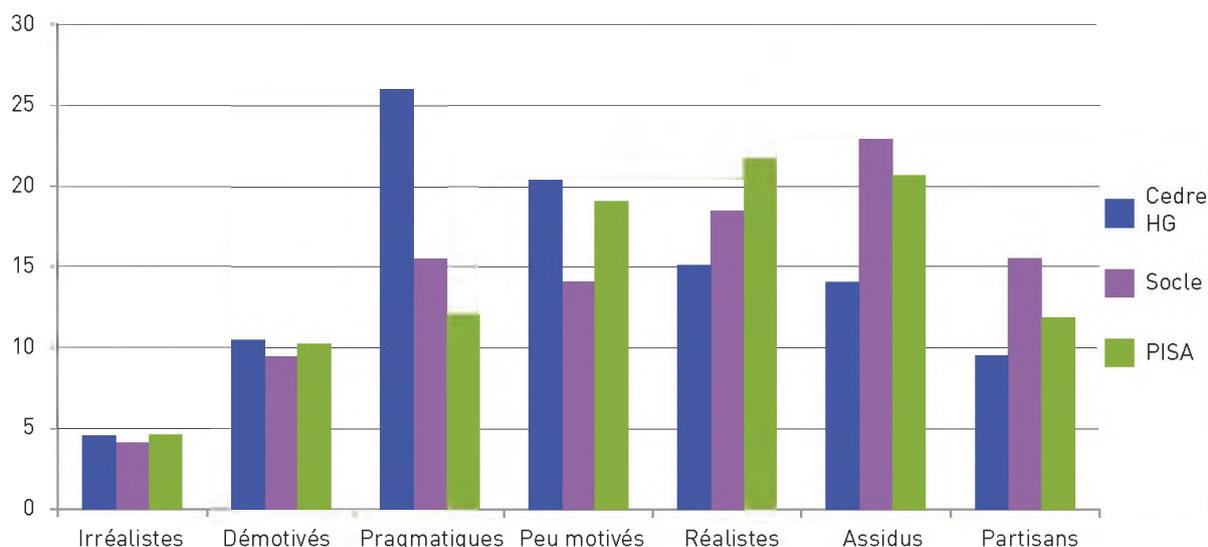
⁵ Évaluation du dispositif d'enseignement intégré de sciences et technologie (EIST). Il s'agit d'une étude longitudinale : les élèves profitant du dispositif et les élèves témoins ont été évalués à cinq reprises entre novembre 2008 et mai 2012 [LE CAM et COSNEFROY, ce numéro, p. 283].

► **Figure 5** Profils de motivation

commun⁶ et les « réalistes » dans PISA ► **Tableau 2**. Les différences entre les évaluations peuvent s'expliquer, en partie, par le terme utilisé dans l'instrument de mesure de motivation. Les évaluations Cedre histoire-géographie et Socle emploient le terme « application », les autres tests mesurant l'investissement des élèves à l'aide du « thermomètre d'effort ». Une représentation graphique de trois évaluations – Cedre histoire-géographie, socle et PISA – donne une vision plus marquée de ces différences ► **Figure 6**. Observons de plus près ces profils en fonction de quelques caractéristiques d'élèves et d'établissements. Pour ne pas surcharger les tableaux, nous ne considérerons que les évaluations Cedre histoire-géographie, socle et PISA qui constituent de bons exemples selon plusieurs aspects. Cedre est une évaluation disciplinaire destinée à mesurer l'atteinte des objectifs fixés par des programmes scolaires officiels. Les tests du socle visent à évaluer les proportions d'élèves ayant acquis les compétences 1 (la maîtrise de la langue française) et 3 (les principaux éléments de mathématiques et la culture scientifique et technologique)

6. Les évaluations du socle commun des connaissances et des compétences [MICONNET et VOUREC'H, ce numéro, p. 141].

► **Figure 6 Répartition des profils de motivation dans Cedre HG, socle et PISA**



Lecture : dans l'évaluation Cedre histoire-géographie la part des élèves « assidus » est de 14 %, contre 21 % dans PISA.

Champs : France métropolitaine, public et privé sous contrat (Cedre) ; France métropolitaine + DOM, public et privé sous contrat (Socle) ; France métropolitaine + DOM (sauf La Réunion), public et privé sous contrat (PISA).

Sources : MENESR-DEPP, évaluations Cedre et socle ; OCDE-PISA 2006.

qui font partie des sept grandes compétences constituant le bagage minimum censé être acquis par tous les élèves à la fin de la scolarité obligatoire. L'évaluation PISA s'intéresse à une génération d'élèves (élèves de 15 ans) et les évalue non sur des connaissances au sens strict mais sur leurs capacités à mobiliser et à appliquer celles-ci dans des situations variées, parfois éloignées de celles rencontrées dans le cadre scolaire. Ces trois évaluations se distinguent également par le format de questions, les tests du socle ne comportant que des QCM alors que Cedre et PISA demandent souvent aux élèves de construire leur réponse. Au-delà de ces différences, ce sont les variations dans l'instrument de mesure qui nous intéressent ici, PISA comportant le « thermomètre d'effort » et les deux autres évaluations le nouvel instrument modifié.

La répartition des profils de motivation varie selon le secteur de scolarisation. Ainsi, les « démotivés » et les « pragmatiques » sont proportionnellement plus nombreux dans des établissements publics – et surtout dans ceux de l'éducation prioritaire (EP) – que dans des collèges privés ► **Tableau 3**.

L'analyse des profils en fonction du sexe met en évidence une moindre motivation de la part des garçons : les « démotivés » et les « pragmatiques » sont proportionnellement plus nombreux parmi eux que parmi les filles ► **Tableau 4**. Les filles semblent ainsi davantage motivées non seulement lors des évaluations à faibles enjeux, mais aussi lorsque la note obtenue compte pour leur bulletin scolaire. Le retard scolaire a également un impact sur la motivation, les « démotivés » étant plus nombreux parmi les élèves ayant redoublé que parmi ceux « à l'heure »

► **Tableau 5**. On note par ailleurs que les « irréalistes » – les élèves qui déclarent s'être plus appliqués (ou avoir fait plus d'effort) en faisant le test que cela aurait été le cas si la note obtenue au test avait compté pour leur bulletin scolaire – sont également plus représentés parmi les élèves redoublants. En revanche,

► **Tableau 3** Répartition des profils de motivation selon le secteur de scolarisation (en %)

Profils de motivation	Cedre HG			Socle		
	Public hors EP	EP	Privé	Public hors EP	EP	Privé
Irréalistes	4	6	4	4	6	2
Démotivés	11	12	10	10	13	7
Pragmatiques	27	29	21	16	18	12
Peu motivés	19	22	22	14	15	15
Réalistes	15	12	16	18	17	21
Assidus	14	12	16	23	19	26
Partisans	10	8	10	16	13	17

Lecture : selon les résultats de l'évaluation Cedre histoire-géographie (HG), la part des élèves « démotivés » est de 12 % dans les établissements de l'éducation prioritaire (EP).

Champs : France métropolitaine, public et privé sous contrat (Cedre) ; France métropolitaine + DOM, public et privé sous contrat (socle).

Sources : MENESR-DEPP, évaluations Cedre et socle.

► **Tableau 4** Répartition des profils de motivation selon le genre (en %)

Profils de motivation	Cedre HG		Socle		PISA	
	Garçons	Filles	Garçons	Filles	Garçons	Filles
Irréalistes	5	4	5	3	5	4
Démotivés	12	9	12	7	12	9
Pragmatiques	28	24	20	11	16	9
Peu motivés	19	22	16	13	19	19
Réalistes	14	16	16	21	19	24
Assidus	12	16	17	29	17	24
Partisans	11	9	14	17	11	12

Lecture : selon les résultats de l'évaluation Cedre histoire-géographie (HG), la part des élèves « démotivés » est de 12 % parmi les garçons.

Champs : France métropolitaine, public et privé sous contrat (Cedre) ; France métropolitaine + DOM, public et privé sous contrat (socle) ; France métropolitaine + DOM (sauf La Réunion), public et privé sous contrat (PISA).

Sources : MENESR-DEPP, évaluations Cedre et socle ; OCDE-PISA 2006.

on observe relativement moins d'écarts selon le retard scolaire pour les profils « pragmatiques » et « peu motivés » que pour les « démotivés », ce qui amène à supposer que les élèves en retard se sentent moins motivés de manière générale (même dans le cas des épreuves scolaires notées), et pas seulement lors des tests à faibles enjeux. Lorsque l'on s'intéresse au statut socio-économique des élèves⁷ selon leur profil de motivation, on observe que ce sont souvent les « irréalistes » qui proviennent des milieux socio-économiques les moins favorisés ► **Tableau 6 p. 130**. Si, dans le cas de l'évaluation Cedre histoire-géographie, l'indice moyen reflétant le statut socio-économique tend à être d'autant plus élevé que l'élève se manifeste motivé, la

7. Mesuré en termes de l'indice de position sociale (IPS) défini par LE DONNÉ et ROCHER [2010] dans Cedre et en termes de l'indice socio-économique international de statut professionnel (HISEI) dans PISA [OCDE, 2009].

► **Tableau 5 Répartition des profils de motivation selon le retard scolaire (en %)**

Profils de motivation	Cedre HG		Socle		PISA	
	À l'heure	En retard	À l'heure	En retard	À l'heure	En retard
Irréalistes	4	7	3	7	2	8
Démotivés	9	16	7	16	9	12
Pragmatiques	25	31	14	18	11	13
Peu motivés	21	18	14	13	19	19
Réalistes	16	11	20	14	23	19
Assidus	15	9	25	17	23	16
Partisans	10	8	16	14	11	13

Lecture : selon les résultats de l'évaluation Cedre histoire-géographie (HG), la part des élèves « démotivés » est de 16 % parmi les élèves en retard dans leurs cursus scolaire.

Champs : France métropolitaine, public et privé sous contrat (Cedre) ; France métropolitaine + DOM, public et privé sous contrat (socle) ; France métropolitaine + DOM (sauf La Réunion), public et privé sous contrat (PISA).

Sources : MENESR-DEPP, évaluations Cedre et socle ; OCDE-PISA 2006.

relation est moins nette pour l'évaluation PISA. Dans PISA, contrairement à Cedre, les « démotivés » et les « pragmatiques » n'ont pas un statut socio-économique moins élevé. Rappelons encore une fois que l'instrument de mesure de motivation n'est pas le même pour les deux évaluations, le terme « application » étant employé dans Cedre et celui « d'effort » dans PISA. On sait que le score des élèves est lié à leur statut socio-économique, les élèves provenant des milieux sociaux moins favorisés obtenant des résultats moins bons [OCDE, 2007]. Il se peut ainsi que les élèves qui déclarent avoir fait peu d'effort pour faire le test – les « démotivés » et les « pragmatiques » – l'aient fait parce que leur niveau est bon et que les exercices leur ont paru faciles. Testons cette hypothèse en observant les scores moyens par profil de motivation.

De manière générale, plus les élèves se déclarent motivés (en termes d'application ou d'effort), plus leur score moyen est élevé ► **Tableau 7**. Néanmoins, ce sont les « assidus » – et non les « partisans » – qui ont le score moyen le plus élevé. Les données PISA montrent que les « démotivés » ont obtenu un score relativement

► **Tableau 6 Statut socio-économique des élèves selon leur profil de motivation**

Profils de motivation	Cedre HG (IPS)		PISA (HISEI)	
	Moyenne	Écart-type	Moyenne	Écart-type
Irréalistes	- 0,21	0,62	50	23
Démotivés	- 0,23	0,70	52	21
Pragmatiques	- 0,14	0,68	53	21
Peu motivés	- 0,06	0,68	51	19
Réalistes	- 0,003	0,66	51	18
Assidus	- 0,02	0,68	52	18
Partisans	- 0,03	0,69	51	18

Lecture : selon les résultats de l'évaluation Cedre histoire-géographie (HG), l'indice IPS moyen des élèves « démotivés » est de - 0,23.

Note : IPS – indice de position sociale défini par LE DONNÉ et ROCHER [2010] ; HISEI – indice socio-économique international de statut professionnel [OCDE, 2009].

Champs : France métropolitaine, public et privé sous contrat (Cedre) ; France métropolitaine + DOM (sauf La Réunion), public et privé sous contrat (PISA).

Sources : MENESR-DEPP, évaluations Cedre ; OCDE-PISA 2006.

élevé. Il s'agit des élèves qui déclarent avoir fourni relativement peu d'effort lors du test et qui disent à la fois que l'effort fourni n'aurait pas été considérablement plus élevé si la note obtenue au test avait compté pour leur bulletin scolaire. Notre hypothèse semble se confirmer : on peut supposer qu'un effort plutôt faible, associé à des scores relativement élevés indique que ces élèves n'ont pas besoin de fournir beaucoup d'efforts car les exercices du test (et les épreuves scolaires notées) sont faciles pour eux.

On peut ainsi constater que la motivation face aux évaluations à faibles enjeux varie selon les caractéristiques d'établissements et d'élèves. Ceci veut dire que lorsque la motivation impacte de manière significative les résultats qu'obtiennent les élèves lors de tels tests, un moindre investissement de certaines sous-populations d'élèves peut introduire des biais dans l'estimation de leur performance et amener à de mauvaises interprétations des résultats. Nous étudions plus en détail la relation entre la motivation et les résultats d'élèves.

► **Tableau 7** Score moyen des élèves selon leur profil de motivation

Profils de motivation	Cedre HG		Socle compétence 1 (français)		Socle compétence 3 (math. et sciences)		PISA (sciences)	
	Moyenne	Écart-type	Moyenne	Écart-type	Moyenne	Écart-type	Moyenne	Écart-type
Irréalistes	213	42	- 0,74	0,95	- 0,78	1,05	425	100
Démotivés	223	52	- 0,54	1,16	- 0,44	1,37	493	116
Pragmatiques	225	42	- 0,33	1,04	- 0,32	1,16	466	99
Peu motivés	246	44	0,10	0,97	0,01	1,00	495	93
Réalistes	256	45	0,32	1,07	0,25	1,07	514	87
Assidus	261	49	0,49	1,05	0,46	1,11	527	90
Partisans	259	55	0,46	1,26	0,46	1,23	517	95

Lecture : selon les résultats de l'évaluation Cedre histoire-géographie (HG), le score moyen des élèves « démotivés » est de 223.

Champs : France métropolitaine, public et privé sous contrat (Cedre) ; France métropolitaine + DOM, public et privé sous contrat (socle) ; France métropolitaine + DOM (sauf La Réunion), public et privé sous contrat (PISA).

Sources : MENESR-DEPP, évaluations Cedre et socle ; OCDE-PISA 2006.

MOTIVATION ET PERFORMANCE

L'évolution du score en fonction de l'évolution de la motivation

Les données longitudinales de l'évaluation EIST nous permettent d'étudier l'évolution des résultats d'élèves d'une session à l'autre. Avec deux sessions d'évaluation (sessions 4 et 5) comportant l'instrument de mesure de l'investissement face au test – en termes d'effort –, nous sommes en mesure de mettre en lien l'évolution du score des élèves avec l'évolution de leur investissement afin d'observer si une augmentation du degré d'investissement s'accompagne d'une évolution positive du score.

En outre, un questionnaire de contexte était adressé aux élèves dans le cadre de cette évaluation. Une partie des questions portait sur leur motivation et leur intérêt vis-à-vis de la discipline évaluée, en l'occurrence les sciences. Sur une échelle à quatre positions, les élèves étaient incités à exprimer leur degré d'accord avec les affirmations suivantes : « *Ce que je fais en science est intéressant* », « *Je participe en science parce que j'aime bien chercher* », « *Je travaille en science parce que j'aime bien cette discipline* », « *Je participe en science parce que j'aime bien faire des expériences* ». Nous avons construit un indicateur synthétisant ces questions à l'aide d'une analyse en composantes principales⁸. Avec cet indicateur de l'intérêt/motivation à l'égard des sciences à notre disposition, nous avons la possibilité de faire la distinction entre les aspects de la motivation spécifiques à la passation du test (*situation-specific motivation*) et les aspects spécifiques au domaine évalué (*domain-specific motivation*).

On peut constater une tendance positive en ce qui concerne l'évolution du score entre la session 4 et la session 5. En revanche, la motivation spécifique à la situation (l'effort fourni au test) ainsi que la motivation spécifique au domaine tendent à diminuer d'une session d'évaluation à l'autre. Lorsque l'on étudie le lien entre ces évolutions, on note qu'elles sont toutes positivement (et significativement) corrélées, mais que cette corrélation n'est pas très élevée ▶ **Tableau 8**. Ainsi, une augmentation de motivation – spécifique à la situation ou au domaine – s'accompagne d'une évolution positive du score. On observe aussi que l'évolution de l'effort fourni au test est un peu plus corrélée à l'évolution du score que l'évolution de l'intérêt vis-à-vis des sciences. Nous avons ensuite construit un modèle de régression dans le but « d'expliquer » le score obtenu lors de la session 5 en fonction de l'évolution de la motivation entre les sessions 4 et 5 ▶ **Tableau 9**. Comme l'indiquent les coefficients de régression, le score obtenu lors de la session 5 est d'autant plus élevé que l'évolution de l'effort fourni au test ainsi que de l'intérêt vis-à-vis des sciences a été positive. En revanche, le pouvoir explicatif du modèle s'avère très faible : l'évolution de ces deux indicateurs de motivation n'explique que 1 % de la variabilité du score obtenu lors de la session 5. Lorsque l'on ajoute au modèle le niveau initial de l'élève – le score obtenu lors de la

▶ **Tableau 8** Relations entre les évolutions du score, de l'effort fourni au test et de l'intérêt vis-à-vis des sciences entre la session 4 et la session 5 (coefficients de corrélation)

	Évolution du score	Évolution de l'effort fourni au test	Évolution de l'indicateur d'intérêt vis-à-vis des sciences
Évolution du score	1	0,148	0,087
Évolution de l'effort fourni au test	0,148	1	0,154
Évolution de l'indicateur d'intérêt vis-à-vis des sciences	0,087	0,154	1

Lecture : l'évolution du score entre les sessions 4 et 5 est liée positivement à l'évolution de l'effort fourni au test (coefficient de corrélation 0,148).

Note : les corrélations significatives au niveau 0,01 sont indiquées en gras.

Champ : France métropolitaine, public.

Sources : MENESR-DEPP, évaluations du dispositif EIST.

⁸. Nous avons effectué une analyse en composantes principales sur les données de la session 4 et nous avons utilisé ces résultats pour calculer les scores factoriels sur les données de la session 5. Seuls les individus participant aux deux sessions ont été retenus pour les analyses présentées ici.

session 4 – les autres coefficients restent significatifs et le coefficient de détermination (R^2) passe à 50 %. Cela nous permet de conclure que c'est surtout le niveau initial de l'élève qui explique son gain en performance entre les deux sessions d'évaluation, l'évolution des aspects motivationnels jouant un rôle considérablement moins important.

► **Tableau 9** Score obtenu lors de la session 5 en fonction de l'évolution des aspects motivationnels (coefficients de régression)

	Modèle 1	Modèle 2
Constante	0,06	- 0,01
Évolution de l'effort fourni au test	0,08	0,11
Évolution de l'indicateur d'intérêt vis-à-vis des sciences	0,08	0,05
Score obtenu lors de la session 4		0,69
R^2	0,01	0,5

Lecture : l'augmentation d'un écart-type du score obtenu lors de la session 4 s'accompagne d'une augmentation du score lors de la session 5 de 0,69 écart-type, l'évolution de l'effort fourni au test et l'évolution de l'indicateur d'intérêt vis-à-vis des sciences étant maintenues constantes.

Note : les coefficients significatifs au niveau 0,01 sont indiqués en gras. R^2 : coefficient de détermination.

Champ : France métropolitaine, public.

Sources : MENESR-DEPP, évaluations du dispositif EIST.

Variabilité inter-classes et intra-classe de la performance en fonction de la motivation

Afin de mieux distinguer l'effet de divers aspects motivationnels et des caractéristiques individuelles et d'établissements sur la variation du score, nous avons employé la méthode de l'analyse multiniveaux (encadré p. 123). Cette méthode permet d'étudier la variabilité, entre établissements et au sein de chaque établissement, de la performance des élèves en fonction de la motivation à répondre au test. Elle permet également de déterminer dans quelle mesure la relation entre la motivation et la performance varie d'un établissement à l'autre. Les analyses présentées ici sont effectuées sur les données de l'évaluation Cedre histoire-géographie. Il s'agit d'un échantillon de classes mais une seule classe par établissement a été tirée au sort, ce qui signifie que ces deux unités sont ici confondues.

Tout d'abord, nous nous sommes intéressés à la décomposition de la variabilité du score en histoire-géographie selon les deux niveaux étudiés (niveau élève et niveau classe). Nous avons construit un modèle « vide » pour connaître la part du score attribuable aux différences entre les établissements et la part attribuable aux différences entre les élèves, au sein d'un même établissement (**modèle 1**)

► **Tableau 10.** En calculant le coefficient de corrélation intra-classe ρ pour ce modèle (encadré p. 137), nous obtenons une valeur de 0,22, ce qui indique que 22 % de la variabilité du score sont dus aux différences entre les classes. Cela signifie que la performance varie plus à l'intérieur des classes qu'entre les classes.

Nous avons ensuite intégré la motivation face au test dans le modèle vide pour observer si cet ajout amène à une réduction de la variance du score entre classes (**modèle 2**).

► **Tableau 10 Modèles multiniveaux**

PARAMÈTRES	MODÈLE 1	MODÈLE 2	MODÈLE 3
	« Vide »	Motivation au test	Pente par établissement
Effets fixes			
Constante	- 0,034 (0,037)	- 0,031 (0,034)	- 0,04 (0,035)
Motivation au test		0,119*** (0,007)	0,119*** (0,008)
Motivation au test si les résultats comptaient pour le bulletin scolaire			
Difficulté perçue du test			
Intérêt vis-à-vis de l'histoire-géographie			
Garçon			
Retard scolaire			
Collège en éducation prioritaire			
Collège privé			
Motivation au test, moyenne par classe			
Effets aléatoires			
Niveau 2 : variance des constantes (variance inter-classes)	0,224*** (0,028)	0,188*** (0,024)	0,192*** (0,024)
Covariance constantes-pentes			0,016*** (0,004)
Variance des pentes			0,003*** (0,001)
Niveau 1 : variance inter-élèves	0,792*** (0,019)	0,734*** (0,017)	0,72*** (0,017)
AIC	10209,3	9906,3	9879,8
BIC	10219,1	9919,3	9899,4

Lecture : toutes les autres variables comprises dans le modèle 8 étant tenues constantes, un garçon obtient un score de 0,11 écart-type plus élevé qu'une fille.

Note : les indicateurs concernant la motivation à répondre au test et la difficulté perçue du test sont centrés par rapport à la moyenne générale. Les unités de ces variables sont exprimées en points. Le score en histoire-géographie ainsi que l'indicateur de l'intérêt vis-à-vis de l'histoire-géographie sont standardisés et ont l'écart-type comme unité de mesure. Les autres variables explicatives n'ont pas d'unité de mesure mais une catégorie de référence (les garçons par rapport aux filles, les élèves en retard par rapport aux élèves « à l'heure », etc.). Les erreurs standards associées aux effets sont présentées entre parenthèses.

*** désigne un effet significatif au niveau 0,01 ; ** désigne un effet significatif au niveau 0,05.

AIC : critère d'information d'Akaike. **BIC :** critère d'information bayésien.

Champ : France métropolitaine, public et privé sous contrat.

Source : MENESR-DEPP, évaluation cedre HG.

MODÈLE 4	MODÈLE 5	MODÈLE 6	MODÈLE 7	MODÈLE 8
Motivation si impact sur bulletin scolaire	Difficulté perçue du test	Intérêt vis-à-vis de l'histoire-géographie	Caractéristiques sociodémographiques et scolaires	Caractéristiques des établissements
- 0,039 (0,034)	- 0,035 (0,032)	- 0,032 (0,031)	0,022 (0,032)	0,011 (0,034)
0,108*** (0,008)	0,099*** (0,008)	0,08*** (0,008)	0,077*** (0,007)	0,07*** (0,007)
0,055*** (0,008)	0,059*** (0,008)	0,047*** (0,008)	0,037*** (0,008)	0,037*** (0,008)
	- 0,096*** (0,007)	- 0,079*** (0,007)	- 0,074*** (0,007)	- 0,074*** (0,007)
		0,205*** (0,015)	0,196*** (0,014)	0,195*** (0,014)
			0,115*** (0,026)	0,11*** (0,026)
			- 0,527*** (0,033)	- 0,526*** (0,033)
				- 0,144** (0,07)
				0,227*** (0,064)
				0,138*** (0,031)
0,186*** (0,023)	0,162*** (0,021)	0,143*** (0,019)	0,125*** (0,017)	0,093*** (0,013)
0,016*** (0,004)	0,014*** (0,003)	0,012*** (0,003)	0,011*** (0,003)	0,01*** (0,003)
0,003*** (0,001)	0,003*** (0,001)	0,003*** (0,001)	0,002*** (0,001)	0,002*** (0,001)
0,712*** (0,017)	0,679*** (0,016)	0,646*** (0,016)	0,607*** (0,015)	0,607*** (0,015)
9839,9	9650,5	9460,6	9208,6	9168,4
9862,7	9676,5	9489,9	9244,4	9214

Autrement dit, les différences dans le degré de motivation expliquent-elles une partie de la variation du score d'un établissement à l'autre ? L'ajout de la motivation au test dans le modèle conduit à une réduction de la variabilité des scores, aux deux niveaux (classe et élève) : la variance inter-classes passe de 0,224 à 0,188 (soit une diminution de 16 %) et la variance inter-élèves passe de 0,792 à 0,734 (soit une diminution de 7 %). Les indices de sélection de modèle (AIC, BIC) diminuent aussi, ce qui indique que le pouvoir explicatif du modèle augmente. Le modèle 2 rend ainsi mieux compte de la structure des données que le modèle 1. On observe également que l'augmentation d'une unité sur l'échelle de la motivation s'accompagne d'un gain de score d'un dixième d'écart-type et que cet effet est significatif.

Dans l'étape suivante, on permet au modèle d'ajuster une pente par établissement afin de tester si la relation entre le score et la motivation à répondre au test varie en fonction de l'établissement (**modèle 3**). En effet, on observe une relation positive entre le score moyen de la classe et la pente entre le score et la motivation : plus le niveau de la classe est élevé, plus la corrélation entre la motivation et le score est forte. Comme les effets aléatoires se sont avérés significatifs, nous les avons gardés pour les modèles suivants.

Nous avons progressivement ajouté des effets fixes : la motivation de l'élève si la note au test avait compté pour son bulletin scolaire (**modèle 4**), la difficulté perçue du test (**modèle 5**) et l'indicateur de l'intérêt vis-à-vis de l'histoire-géographie (**modèle 6**). Le questionnaire de contexte adressé aux élèves à la fin de leur cahier d'évaluation comportait un certain nombre d'items relatifs à l'intérêt vis-à-vis de l'histoire-géographie. Sur une échelle à cinq positions, les élèves étaient incités à indiquer à quelle fréquence ils s'intéressent, en dehors du collège, à l'histoire, à la géographie, à l'éducation civique « *en lisant des livres ou des revues sur ces sujets* », « *en regardant des émissions de télévision ou des films* », « *en cherchant des documents sur Internet* », « *en visitant des musées, des lieux historiques, des sites géographiques* » [jamais, rarement, de temps en temps, souvent, très souvent]. Le questionnaire les interrogeait également sur leurs sentiments vis-à-vis de la discipline : « *À propos de l'histoire, la géographie, l'éducation civique, vous diriez...* » [j'adore, j'aime bien, j'aime moyennement, je n'aime pas du tout, je déteste]. Nous avons employé l'analyse en composantes principales afin de construire un indicateur synthétisant tous ces items.

Ensuite, nous avons intégré dans le modèle les caractéristiques sociodémographiques et scolaires des élèves telles que le sexe et le retard scolaire (**modèle 7**) et, enfin, le secteur et le degré moyen de motivation caractérisant les établissements (**modèle 8**). On observe que les indices AIC et le BIC diminuent à chaque étape, ce qui met en évidence une augmentation progressive du pouvoir explicatif du modèle. Par rapport au modèle vide, le modèle complet (**modèle 8**) amène à une réduction de la variance inter-classes de 59 % (de 0,224 à 0,093) et de la variance inter-élèves de 23 % (de 0,792 à 0,607).

On note que l'augmentation d'une unité sur l'échelle de la motivation au test s'accompagne d'un gain de 0,07 écart-type de score, toutes choses égales par ailleurs (modèle 8). Une unité supplémentaire sur l'échelle de la motivation, si l'épreuve était notée, amène quant à elle à une augmentation de score de 0,04 écart-type. La difficulté perçue du test est liée négativement au score : un point supplémentaire sur cette échelle entraîne une réduction du score de 0,07 écart-type. L'augmentation, d'un écart-type, de l'indicateur de l'intérêt/motivation vis-à-vis-de l'histoire-géographie est associé à un gain de score de 0,2 écart-type. Toutes choses égales par ailleurs, un garçon obtient un score d'un dixième d'écart-type plus élevé qu'une fille et un redoublant un score d'un demi écart-type moins élevé qu'un élève « à l'heure ». Enfin, le score varie également en fonction des caractéristiques d'établissements : toutes choses égales par ailleurs, un élève scolarisé dans le secteur de l'éducation prioritaire obtient un score en histoire-géographie moins élevé (de 0,14 écart-type) et celui inscrit dans un collège privé un score plus élevé (de 0,23 écart-type) par rapport à celui suivant sa scolarité dans un établissement public hors éducation prioritaire. Outre le niveau individuel de motivation à répondre au test, le niveau moyen

MODÈLES MULTINIVEAUX

L'analyse multiniveaux, comme le nom l'indique, prend en compte la structure hiérarchique des données, en l'occurrence organisée sur deux niveaux : élève et établissement (ou bien la classe). L'emploi des modèles multiniveaux permet ainsi d'évaluer la variabilité des résultats à l'intérieur des établissements [entre les élèves] ainsi qu'entre les établissements.

Dans l'analyse multiniveaux, on peut distinguer les effets fixes du modèle, qui sont les coefficients de régression, et les effets aléatoires du modèle, c'est-à-dire les estimations de variance. Le modèle « vide » consiste en une décomposition de variance :

$$y_{ij} = \beta_{0j} + \varepsilon_{ij}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

où y_{ij} est le score de l'élève i dans l'établissement j , β_{0j} est la constante de l'établissement j , ε_{ij} le résidu pour l'élève i dans l'établissement j , γ_{00} la constante générale et u_{0j} l'écart entre la constante générale et la constante de l'établissement j [parfois appelé « effet établissement »]. Le coefficient de corrélation intra-classe ρ montre la part de la variance totale du score qui peut être expliquée par les différences entre les établissements :

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_{\varepsilon0}^2}$$

où σ_{u0}^2 est la variance inter-établissement – soit $\text{Var}(u_{0j})$ – et $\sigma_{\varepsilon0}^2$ est la variance inter-élèves, soit $\text{Var}(\varepsilon_{ij})$.

Avec l'introduction d'une variable explicative – par

exemple, la motivation à répondre au test X – comme un effet fixe, l'équation peut être écrite :

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

Ce modèle a deux composantes aléatoires, la variance de ε_{ij} et la variance de u_{0j} , et deux paramètres fixes, γ_{00} et γ_{10} (i.e. le coefficient de régression de la variable X).

Nous avons considéré, plus haut, que l'effet de la variable indépendante sur la variable dépendante est fixe, c'est-à-dire que la relation entre la motivation à répondre au test et la performance est la même dans chaque établissement. Or, les établissements peuvent se distinguer au regard de ce lien. L'analyse multiniveaux permet de prendre en compte ces différences en définissant l'effet de la variable explicative comme un effet aléatoire, autrement dit, en ajustant une pente de régression par établissement. L'équation avec pentes aléatoires peut s'écrire :

$$y_{ij} = \alpha_j + \beta_{1j}X_{ij} + \varepsilon_{ij}$$

$$\alpha_j = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Dans ce modèle, le coefficient de régression de la variable X a deux composantes : un coefficient de régression général identique pour tous les établissements (la part fixe, noté γ_{10}) et un coefficient relatif à l'établissement correspondant (la partie aléatoire, noté u_{1j}).

Pour en savoir plus sur les modèles multiniveaux, le lecteur est invité à consulter l'ouvrage de GOLDSTEIN [2003].

de motivation par classe joue un rôle sur le score, les élèves provenant des classes plus motivées en moyenne ayant des scores plus élevés, toutes choses égales par ailleurs (l'augmentation d'une unité de cet indicateur par classe est associée à un gain de score de 0,14 écart-type)¹¹. Ce dernier résultat est sans doute à relier aux effets des conditions de passation, dont on sait qu'elles peuvent varier selon les établissements ► **Encadré ci-dessus**. L'influence sur la motivation des élèves du personnel de l'établissement en charge de l'évaluation doit être davantage explorée lors de futures études.

11. Afin de vérifier si de tels résultats sont spécifiques à cette évaluation, nous avons testé le même modèle sur les données de l'évaluation du socle commun. Nous observons quelques variations mais le profil de régression est identique, nous retrouvons les mêmes tendances et les mêmes ordres de grandeur des coefficients.

CONCLUSIONS, PRÉCONISATIONS

L'analyse et la comparaison des données recueillies par des instruments de mesure de motivation – « thermomètre d'effort » ou « échelles d'application » – mettent en évidence des différences dans les résultats ainsi que dans les conclusions qui peuvent en être tirées. La motivation des élèves étant liée à la difficulté (perçue) du test, la nécessité d'ajouter une échelle pour mesurer cette difficulté se manifeste. Les échelles « d'application » semblent ainsi mieux adaptées pour rendre compte du degré de motivation des élèves français que le thermomètre « d'effort » utilisé dans PISA. En outre, comportant une charge de lecture considérablement moindre que ce dernier, l'instrument modifié est censé moins désavantager les lecteurs faibles.

La motivation face au test, comme nous l'avons vu, varie selon les caractéristiques d'élèves, les garçons et les redoublants se manifestant relativement moins impliqués. De même, la motivation semble moindre dans le secteur de l'éducation prioritaire, ce qui confirme, dans le contexte français, les tendances repérées dans une étude précédente [JAKWERTH, STANCAVAGE, REED, 1999] selon laquelle la motivation posait problème dans des établissements caractérisés par le plus faible niveau des élèves.

La relation entre la motivation à répondre au test et la performance semble, selon notre étude, plutôt modeste. En outre, les différences en termes de degré de motivation ne sont pas considérables entre les établissements. Cependant, pour mieux explorer la question de motivation des élèves face à un test sans enjeux pour eux, nous avons mené une expérience à partir de l'évaluation Cedre mathématiques en mai 2014. Cette expérience – en cours d'analyse – vise à comparer les performances de deux groupes d'élèves : un groupe expérimental ayant au préalable reçu l'information que l'épreuve sera notée, et un groupe témoin qui passe le test dans des conditions habituelles.

BIBLIOGRAPHIE

- BOBINEAU M., 2013, *Évaluations dites « à faibles enjeux » : quelle perception et implication de la part des élèves ? Étude qualitative à partir de CEDRE Sciences 2013*, rapport de stage, MENESR-DEPP.
- BUTLER J., ADAMS R. J., 2007, "The Impact of Differential Investment of Student Effort on the Outcomes of International Studies", *Journal of Applied Measurement*, vol. 8, n° 3, p. 279-304.
- DECI E. L., RYAN R. M., 1985, *Intrinsic motivation and self-determination in human behavior*, New York, Plenum.
- EKLÖF H., 2008, "Test-taking motivation on low-stakes tests: A Swedish TIMSS 2003 example", *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, vol. 1, p. 9-21.
- GOLDSTEIN H., 2003, *Multilevel Statistical Models*, 3^e éd., Londres, Edward Arnold.
- JAKWERTH P. M., STANCAVAGE F. B., REED E., 1999, *An Investigation of Why Students Do Not Respond to Questions*, NAEP Validity Studies (NVS) Panel, Palo Alto, American Institute for Research.
- KESKPAIK S., ROCHER T., 2012, « Les évaluations à faibles enjeux : quel rôle joue la motivation ? Une expérience à partir de PISA », communication dans le cadre du 24^e colloque de l'ADMEE-Europe, Luxembourg.
- LE DONNÉ N., ROCHER T., 2010, « Une meilleure mesure du contexte socio-éducatif des élèves et des écoles. Construction d'un indice de position sociale à partir des professions des parents », *Éducation & formations*, n° 79, MENJVA-DEPP, p. 103-115.
- NDINGA P., FRENETTE E., 2010, « Élaboration et validation de l'Échelle de motivation à bien réussir un test (EMRT) », *Mesure et évaluation en éducation*, vol. 33, n° 3, p. 99-123.
- OCDE, 2009, *PISA 2006, Technical Report*, Paris, OCDE.
- OCDE, 2009, *PISA Data Analysis Manual*, 2^e éd., Paris, OCDE.
- OCDE, 2007, *PISA 2006, Les compétences en sciences, un atout pour réussir, vol. 2 : Données*, Paris, OCDE.
- O'NEIL H. F., ABEDI J., LEE C., MIYOSHI J., MASTERGEORGE A., 2004, *Monetary Incentives for Low-Stakes Tests*, CSE Report 625, Los Angeles, CRESST/UCLA.
- PENK C., POEHLMANN C., ROPPELT A., 2013, "Do Test-Takers Give Their Best? Motivational Determinants of Test Performance in Low-Stakes Assessments", article présenté au congrès annuel de l'American Educational Research Association, San Francisco.

