

June 2015

## Diagnostic des conceptions en sciences susceptibles d'expliquer les différences de performances à une évaluation internationale entre le Québec et le Maroc

Guy Norbert Loubaki

Université du Québec à Montréal, Montréal, loubaki2@yahoo.fr

Patrice Potvin

Université du Québec à Montréal, Montréal, potvin.patrice@uqam.ca

Lalla Rabia Hijazi

École Normale Supérieure Takaddoum Rabat, Rabat, rabia.hijazi@gmail.com

Jesùs Vázquez-Abad

Université de Montréal, Montréal, j.vazquez-abad@umontreal.ca

Follow this and additional works at: <http://ir.lib.uwo.ca/cie-eci>

---

### Recommended Citation

Loubaki, Guy Norbert; Potvin, Patrice; Hijazi, Lalla Rabia; and Vázquez-Abad, Jesús (2015) "Diagnostic des conceptions en sciences susceptibles d'expliquer les différences de performances à une évaluation internationale entre le Québec et le Maroc," *Comparative and International Education / Éducation Comparée et Internationale*: Vol. 44: Iss. 1, Article 3.

Available at: <http://ir.lib.uwo.ca/cie-eci/vol44/iss1/3>

This Research paper/Rapport de recherche is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Comparative and International Education / Éducation Comparée et Internationale by an authorized administrator of Scholarship@Western. For more information, please contact [jpater22@uwo.ca](mailto:jpater22@uwo.ca).

**Diagnostic des conceptions en sciences susceptibles d'expliquer les différences de performances à une évaluation internationale entre le Québec et le Maroc**  
**Diagnostic of conceptions in the sciences that could explain the differences in performance in an international assessment in Quebec and Morocco**

Guy Norbert Loubaki, Université du Québec à Montréal  
Patrice Potvin, Université du Québec à Montréal,  
Lalla Rabia Hijazi, École Normale Supérieure Takaddoum Rabat  
Jesùs Vázquez-Abad, Université de Montréal

**Résumé**

Cet article présente un aperçu des résultats d'une étude doctorale qui s'est inscrite dans une perspective didactique d'explication des différences de performance à une évaluation internationale pour des sujets habitants des contextes géographiques où les habitudes de vie diffèrent considérablement. Dans une démarche de diagnostic de conceptions initiales en sciences, cette étude a poursuivi l'objectif de mieux comprendre et d'expliquer les différences de performance réalisées par des jeunes marocains et québécois de 15 ans dans une évaluation internationale combinant les questions en sciences des enquêtes TEIMS 2007 et PISA 2006. Les résultats de cette étude ont permis à la didactique d'esquisser des hypothèses de recherche portant sur l'équité dans l'explication des différences de performance entre élèves de contextes géographiques considérablement différents et de nuancer la valeur des évaluations internationales et des interprétations qui peuvent découler.

**Abstract**

This article presents an overview of the results of a doctoral study explaining the differences in performance in an international assessment of participants living in geographical contexts where lifestyles differ greatly. This study's objective is to better understand and explain the differences in performance obtained by 15-year-old Moroccan and Quebec youth in an international assessment combining sciences questions from studies on TIMSS 2007 and PISA 2006. The study results will eventually propose research hypotheses on equity explaining the differences in performance between students from significantly different geographical contexts, and will nuance the value of international assessments and the interpretations that may arise from them.

**Mots clés:** diagnostic des conceptions initiales, science, différence de performance, contexte géographique, évaluation internationale

**Keywords:** initial conceptions in sciences; cultural context; international assessment

**Introduction**

La présente étude s'inscrit dans une perspective didactique de diagnostic des conceptions en sciences (Tsai & Chou, 2002; Kraus & Minstrell, 2002). Notre problématique émerge des constats (Kjaernsli, Angell & Lie, 2002; Angell, 2004; Turmo, 2003; Allal, 2007; Wu, 2008) liés aux difficultés d'expliquer les différences de performance à une évaluation internationale pour des sujets habitants des contextes géographiques où les habitudes de vie diffèrent considérablement. Cette étude poursuit l'objectif de mieux comprendre et d'expliquer les différences de performance réalisées par des jeunes marocains et québécois de 15 ans dans une évaluation internationale de type TEIMS. Elle s'opère dans une approche méthodologique mixte à dominance qualitative avec une posture interprétative

qui vise à comprendre les conceptions non scientifiques. Ce travail est présenté en deux parties. La première partie présente les bases théoriques de l'étude notamment le contexte, les questions, les objectifs et le cadre théorique. La deuxième partie présente la méthodologie, les résultats, l'analyse et l'interprétation.

### **Contexte d'étude**

Les enquêtes internationales en éducation évaluent auprès des jeunes soit des compétences générales pour vivre et travailler dans une société post-industrielle (PISA) soit des connaissances à propos des acquis scolaires (TEIMS). Ces enquêtes proposent essentiellement de fournir des indicateurs de pilotage aux responsables des systèmes éducatifs des pays membres. Les opérations menées par ces institutions sur le terrain s'inscrivent entre autres dans un souci de connaissance et de jugement de valeur sur la qualité des systèmes d'éducation de chaque pays (Bulles, 2010). Ce dernier cas signifie que certaines valeurs éducatives seraient reconnues de façon universelle, valables de la même façon pour l'ensemble des systèmes d'éducation. Ce sont ces valeurs éducatives dites standards internationaux (Meunier, 2005) qui permettent la comparabilité et le classement des différents pays sur une échelle de score de performance.

Cette étude se réalise donc dans un contexte où les scores de performance des pays aux enquêtes internationales en éducation (PISA, TEIMS, SACMEQ, PASEC, etc.) sont sujets à diverses interprétations (Grisay, de Jong, Gebhardt, Berezner, & Halleux-Monseur, 2007) occasionnant parfois des vives critiques de formats ou de contenus d'items proposés (Harlow, & Jones, 2004). Les interprétations divergent selon le rôle et la place attribués au contexte culturel. Ce contexte conditionne la construction des connaissances et les réponses données par l'élève (Angell, Kjærnsli, & Lie, 2006; Grønmo, Kjærnsli, & Lie, 2004).

Dans la littérature portant sur les grandes enquêtes internationales TEIMS (Martin, Mullis, Foy, & Stanco, 2012, 2013; Martin, & Mullis, 2013), les différences de score de performance souvent récurrentes à travers divers tests, tel que l'indique le tableau 1 ci-après, n'ont pas toujours trouvé d'explication du point de vue qualitatif (Goldstein, 2004) ou didactique (l'objet de cette étude). Le choix des pays s'est fait dans le souci d'illustrer les différences de score de performance par item en fonction des divers contextes géographiques.

Les données de ce tableau indiquent que malgré l'éloignement des contextes géographiques et la différence des habitudes de vie, les performances de l'Angleterre, du Japon et du Québec par item sont souvent très proches tandis que celles du Maroc s'éloignent considérablement. Ce démarquage de performance laisse penser que les élèves québécois et marocains, par exemple, ont certainement diverses manières de conceptualiser les phénomènes scientifiques selon leurs habitudes de vie quotidiennes. Ce constat a déjà été évoqué par Allal (2007), qui soutient que les tâches qui sont exécutées au sein des communautés de pratiques placent les élèves à une inégale distance des «tâches d'évaluation». Cette distanciation engendre des difficultés à établir l'équité sinon à expliquer les différences de performances observées. Allal soutient que les élèves de contextes différents ont nécessairement acquis leurs savoirs et compétences grâce à leur insertion dans des communautés de pratiques extrêmement variables (Allal, 2007). Ce constat amène à préciser le problème didactique sous-jacent ces différences de performance.

*Tableau 1 : résultats des enquêtes TEIMS 2003-2007-2011 en science*

Items en science grade 8	pays	Performance en % de réussite		
		2003	2007	2011
1	Québec	45**	36	87
	Japon	10	65	95
	Angleterre	47	28	88
	Maroc	1*	15	80
2	Québec	47**	11	84
	Japon	47	71	99
	Angleterre	44	48	92
	Maroc	16*	15	82
3	Québec	50**	52	76**
	Japon	58	68	82
	Angleterre	48	66	69
	Maroc	6*	38	42*
4	Québec	41	38**	80**
	Japon	31	75	71
	Angleterre	57	57	79
	Maroc	16	8*	44*
5	Québec	79**	56	39**
	Japon	77	75	72
	Angleterre	74	53	61
	Maroc	33*	66	7*
6	Québec	86**	73**	65**
	Japon	92	82	50
	Angleterre	78	77	65
	Maroc	6*	44*	33*
7	Québec	89	89	57**
	Japon	76	82	52
	Angleterre	88	85	56
	Maroc	66	60	10*
8	Québec	87	79	44**
	Japon	93	89	30
	Angleterre	95	91	59
	Maroc	81	51	4*

(Source IEA Eighth Grade Results for the TIMSS International Benchmarks in Science 2003, pp. 74-87; 2007, pp. 97-110; 2011, pp. 120-136)

### **Problème d'étude**

Les différences de performance aux enquêtes internationales, laissent penser que malgré que les tâches d'évaluation internationales soient décontextualisées, il existerait des conceptions qui pourraient orienter les réponses des élèves lors de la réalisation de ces tâches. Cependant, le hasard peut également faire partie de ces performances. L'un des problèmes qui émerge des constats précédents est celui de pouvoir trouver des explications satisfaisantes au sujet des différences de performance des élèves de contextes géographiques présentant des habitudes de vie différentes. En introduisant l'estimation du degré de certitude à la réponse (Caleon, & Subramaniam, 2010) et sa justification par

l'élève nous entendons faire émerger les conceptions qui motivent le choix d'une réponse.

Plusieurs études didactiques (Tsai, & Chou, 2002; Simard, Harvey, & Samson, 2014; Alci, 2015) ont révélé l'influence des conceptions dans la réalisation des tâches d'apprentissage et d'évaluation. Pour la présente étude, les conceptions des élèves et leur degré de certitude à la réponse deviennent des objets de recherche susceptibles de nourrir les explications de différence de performance observées.

Notre problème de recherche en didactique des sciences est donc de comprendre et d'expliquer qualitativement les différences de performances aux enquêtes internationales à partir des conceptions en sciences des élèves de contexte géographique de vie considérablement différents. Plus particulièrement, cette étude s'intéresse aux difficultés qu'éprouvent ces élèves à pouvoir réaliser de façon équitable des tâches demandées. De ce problème découle des questionnements et des objectifs ci-après.

### **Question et objectif de recherche**

Au regard du problème d'étude qui se résume à la recherche de l'équité dans la compréhension et la réalisation des tâches dites décontextualisées, une question globale se pose: Quelles sont les conceptions en sciences pouvant être à l'origine des réponses d'élèves et qui sont susceptibles d'expliquer les différences de performance qu'on observe lors de la réalisation d'une tâche d'évaluation internationale?

Si cette étude réussit à trouver des éléments de réponses à cette question, alors elle permettra à la didactique des sciences d'atteindre l'objectif général suivant: Mieux comprendre et expliquer les différences observées sur les résultats d'une évaluation internationale à partir des conceptions en sciences des élèves de contextes géographiques présentant des habitudes de vie considérablement différentes.

Pour opérationnaliser cet objectif général, des objectifs spécifiques ci-après sont nécessaire:

- (1) Identifier des items d'une tâche d'évaluation internationale en sciences qui présentent les plus importantes différences de performance entre deux contextes géographiques présentant des habitudes de vie considérablement différentes.
- (2) Identifier les conceptions en sciences qui sont derrière ces réponses et qui sont susceptibles d'être à l'origine des différences de performance observées.

Pour ce faire, un cadre théorique de traitement des résultats mérite d'être précisé.

### **Cadre théorique**

Dans cette étude, le concept de conception (Giordan, & De Vecchi, 1987) est défini et utilisé au sens que lui donne Balacheff (1995) contrairement au concept de représentation (Orange, & Ravachol, 2013) qui elle est polysémique. Cette étude s'inspire donc de la perspective de Balacheff (1995) qui établit des balises en quadruplet P, R, L,  $\Sigma$  permettant d'identifier des conceptions en sciences. Ces balises proviennent de l'acception du concept de conception issue du paradigme de l'erreur et de la problématique des obstacles (Bachelard, 1938).

Balacheff postule à propos de la dynamique du système sujet/milieu que la connaissance humaine est constituée d'une multiplicité de «conceptions» localement valides. Chaque «conception» existe, dit l'auteur, si elle est opérationnelle dans un certain domaine de validité c'est-à-dire qu'elle permet de résoudre un ensemble de problèmes. Balacheff caractérise une conception C, par un quadruplet P, R, L,  $\Sigma$  dans lequel: -P est un

ensemble de problèmes sur lequel C est opératoire; - R est un ensemble d'opérateurs; - L est un système de représentation, il permet d'exprimer les éléments de P et de R; -  $\Sigma$  est une structure de contrôle, elle assure la non-contradiction de C. Ce sont ces balises qui vont nous permettre d'analyser et d'interpréter les résultats de cette étude. Des objectifs de recherche et du cadre théorique découlent la méthodologie ci-après.

### **Méthodologie**

Pour comprendre et expliquer les différences de performance, nous ferons émerger des conceptions qui orientent le choix des réponses à partir de deux données: la justification de la réponse et le degré de certitude exprimée. La prise en compte de l'estimation de degré de certitude exige une quantification des résultats avant de les expliquer qualitativement. Cette posture bien qu'interprétative nécessite une approche à la fois quantitative et qualitative. C'est donc une approche mixte à dominance qualitative suivant une posture interprétative. Cette méthodologie s'adresse à une population dont il convient de préciser les caractéristiques.

### ***Caractéristiques de la population, échantillonnage et échantillon***

La population à l'étude est constituée de l'ensemble des élèves de 15 ans qui fréquentent les écoles secondaires au Maroc et au Québec. Le choix de l'âge vient du fait qu'à partir de 15 ans l'élève est supposé être à la fin du cycle scolaire obligatoire. Il est donc capable de faire preuve d'un minimum de connaissances requises pour vivre et travailler dans une société post-moderne. C'est le standard international fixé par PISA (OCDE, 2010). Le choix de ce standard permet de minimiser le problème d'harmonisation des curriculums entre États. Même-ci TEIMS sélectionne les élèves d'un même niveau, le contenu notionnel étudié au même niveau n'est pas toujours le même selon les différents systèmes éducatifs (Mons, 2008).

Le choix du Maroc et du Québec est dicté par le fait que les deux participent aux enquêtes TEIMS, mais également par la différence des habitudes de vie quotidienne et par les importants écarts de performance répétitifs de 2003 à 2011. Pour réaliser cette étude, un échantillon est choisi. Le plan d'échantillonnage de cette étude est basé sur la technique d'échantillonnage non aléatoire par quotas (Kaplan, 2004). Cette technique permet de constituer des quotas des élèves de différentes écoles stratifiées. La taille de l'échantillon retenu pour cette étude est de 120 élèves par groupe de même niveau scolaire pris au hasard dans différentes écoles. Ce nombre bien qu'insignifiant en termes de généralisation permet néanmoins qualitativement de faire émerger les tendances des différences observables. Pour ce faire, des instruments et des techniques de collecte des données sont nécessaires.

### **Instruments et techniques de collecte des données**

C'est dans le but de recueillir les conceptions de l'élève sur sa propre réponse que nous avons choisi de refaire un test. Nous avons réutilisé 20 questions issues du questionnaire TEIMS 2007 et PISA 2006.

### ***Pour atteindre le premier objectif spécifique***

Les 20 questions ont été choisies en fonction des catégories de problèmes tels que présentés dans le tableau ci-après.

**Tableau 2: choix des questions en fonction des catégories de problèmes**

Question	Catégorie des problèmes
Question 1°; 2°; 3	Comprendre un phénomène scientifique à partir d'un usage technologique
Question 4°;5°; 8°; 13°; 14	Comprendre un phénomène scientifique à partir d'une pratique quotidienne
Question 11°; 15°; 17°; 18°; 20	Comprendre un phénomène scientifique à partir d'une expérience de laboratoire
Question 6°; 7°; 9°; 10°; 12°; 16°; 19	Comprendre un phénomène scientifique à partir des résultats d'une recherche.

Ces questions à choix multiple, à l'instar des travaux de Chen, Chang, & Guo (2004) et de Park & Bolt (2008), ont été complétées par l'estimation de degré de certitude sur une échelle de likert (Caleon & Subramaniam, 2010). Ce questionnaire spécifique par sa forme et son contenu a été administré aux 120 élèves. Le tableau ci-dessous a permis de traiter les résultats obtenus.

**Tableau 3 : Échelle de classification et de catégorisation des réponses**

Certitude (C)		C1	C2	C3	C4	C5
		Pas du tout sûr	Pas très sûr	Légèrement sûr	Très sûr	Extrêmement sûr
Réponse (R)	Si fausse	N1	N2	N3	N4	N5
	Si juste	N'1	N'2	N'3	N'4	N'5

Les lettres N indiquent les effectifs des réponses par certitude exprimée. Les différents choix de réponses sont représentés par une légende.

L'exploitation de cette échelle est inspirée des travaux de Gilles (2002). Cet auteur présente une technique d'articulation de la réponse R au degré de certitude C comme suit: 1) Si l'élève exprime une certitude très faible «pas du tout sûr» alors que sa réponse est incorrecte ; 2) Si l'élève exprime une certitude faible, donc il reconnaît qu'il n'est pas sûr de sa réponse et que la réponse est incorrecte; 3) Si l'élève exprime une certitude faible, donc il reconnaît qu'il n'est pas sûr de sa réponse et que la réponse est correcte; 4) Si l'élève n'exprime pas de certitude et ne donne pas de réponse; 5) Si l'élève exprime une certitude élevée et que sa réponse est correcte.

Cependant, cette technique est confrontée aux limites dues aux données manquantes relatives à la non-réponse ou à la certitude non exprimée. Gilles (2002) a élaboré un procédé qui permet d'identifier et de gérer ces limites.

Ce procédé obéit aux principes suivants:

1) L'élève n'écrit rien (il ne donne pas de réponse et n'exprime pas sa certitude). Ceci est considéré comme une omission; 2) L'élève n'écrit pas de réponse mais exprime une

certitude. Dès qu'il n'y a pas de réponse, c'est considéré comme une omission; 3) L'élève écrit une réponse mais n'exprime pas sa certitude. La réponse est correcte. La réponse est acceptée, la certitude «pas très sûr» est appliquée par défaut. S'il n'avait aucune confiance en sa réponse et qu'il pensait mettre la certitude «pas très sûr», ça revient au même pour lui; 4) L'élève écrit une réponse mais n'exprime pas la certitude. Si sa réponse est incorrecte, elle sera néanmoins acceptée. La certitude «pas du tout sûr» est appliquée par défaut. Dans ce cas, l'avantage va à l'élève et revient au même pour lui que s'il avait exprimé une certitude «pas du tout».

Cette technique a donc été appliquée comme norme pour gérer les données manquantes. Par conséquent, l'exploitation des résultats a permis de sélectionner 14 items définitivement retenus pour réaliser notre étude qualitative.

### ***Pour atteindre le deuxième objectif spécifique***

Afin de recueillir les conceptions de chaque élève au sujet de sa réponse et de sa certitude, nous avons procédé aux entretiens individuels semi-directifs (Boutin, 2000) compte tenu du respect des balises définies dans le cadre théorique. Les questions ont été posées une à une avec la possibilité de relance lorsqu'il y a ambiguïté entre la réponse, la certitude exprimée et leur justificatif.

### ***Technique d'analyse des résultats d'entretiens***

Les entretiens ont porté essentiellement sur les 14 items qui ont permis d'observer d'importantes différences de performance lors de notre test. Les verbatim obtenus ont été analysés suivant le principe ci-après.

### ***Principe d'élaboration des indicateurs de codage des verbatim***

Six critères de codage ont été retenus. Les quatre premiers sont les caractéristiques P, R, L et  $\Sigma$  qui découlent des balises du cadre théorique. Les deux autres sont : la description du concept principal du problème par l'élève et la véracité de la réponse (vrai/faux). Dans cette étude nous appelons indicateur le mot ou le groupe de mots qui permet d'établir un lien descriptif ou explicatif entre un passage ou un concept pertinent qui ressort des propos de l'élève. Les indicateurs découlent donc d'un va-et-vient (une mise en concordance) de correspondance sémantique entre les six critères et les propos de l'élève. La suite présente les relations entre les critères, les indicateurs de codage des verbatim et leurs rôles :

(1) Le critère «Concept principal» permet d'identifier si l'argument que l'élève soutient pour décrire le problème ou la question avec ses propres mots a un lien avec le concept principal du problème posé. Deux indicateurs permettent donc de vérifier ce lien: «étrange au problème» est dit pour un argument qui n'a aucun lien avec le concept principal du problème posé ; « pertinent au problème » est dit d'un argument qui a un lien direct avec le concept principal du problème posé.

(2) Le critère «Problème P» permet d'identifier si dans les propos de l'élève, ce dernier reformule le problème avec ses propres mots ou s'il restitue le problème tel que présenté dans l'énoncé de l'item. Deux indicateurs permettent de vérifier ce lien: «Ne reformule pas le problème» est dit d'un argument qui restitue le contenu du problème de l'item; «Reformule le problème» est dit d'un argument basé sur d'autres mots pour reformuler le problème.

(3) Le critère «Opérateur R» permet d'identifier si les ressources mobilisées pour

réaliser la tâche découle des circonstances présentées dans l’item ou d’un raisonnement personnel. Trois indicateurs permettent de vérifier ce lien : « Exploite le texte » est dit d’une stratégie basée sur le décodage de l’énoncé de l’item ; « Logique personnelle » est dite d’une stratégie qui est une métaconnaissance ; « Renvoie au texte de la question » est dit d’une stratégie comparable à un plagiat.

(4) Le critère « Représentation L » permet d’identifier l’origine des ressources mobilisées pour réaliser la tâche. Trois indicateurs permettent de vérifier ce lien : « Réfère à une source médiatique » est dit d’un argument provenant des moyens autres que l’école et les pratiques sociales ; « Réfère à une pratique sociale » est dit d’un argument provenant des pratiques de vie quotidienne ; « Réfère au savoir scolaire » est dit d’un argument provenant des savoirs codifiés dans les programmes de formation.

(5) Le critère « Validation  $\Sigma$  » permet d’identifier la conviction avec laquelle une réponse est soutenue. Deux indicateurs permettent de vérifier ce lien : « Exprime une certitude » est dit d’une réponse dont l’élève est convaincu de ses arguments ; « Exprime un doute » est dit d’une réponse dont l’élève n’est pas convaincu de ses arguments.

(6) Le critère « Réponse » permet d’accepter ou de réfuter la véracité d’une réponse. Deux indicateurs permettent de vérifier ce lien : « Bonne réponse » est dite d’un choix de réponse juste. « Mauvaise réponse » est dite d’un choix de réponse faux.

Le guide d’entretiens suivant a été utilisé pour conduire les entretiens conformément au cadre théorique.

**Tableau 4 : Guide d’entretiens**

Critères permettant de caractériser une conception	Question du chercheur	Réponse de l’élève
P (Ensemble de problèmes)	Comment comprends-tu la question posée°? Ou : peux-tu reformuler la question dans tes propres mots°?	
R (Ensemble d’opérateurs pour transformer un problème de P à un autre)	Pourquoi as-tu choisi telle réponse et pas telle autre alors que tu as exprimé telle certitude°?	
L (Systèmes de représentation)	Que penses-tu de telle idée°?	
$\Sigma$ (Critères de validation)	Qu’est ce qui te rassure que ta réponse est vraie°? Ou : Peux-tu me donner des exemples ou des arguments de choses que tu sais et qui appuient ta réponse°?	

### Résultats, analyse et interprétation

Les résultats sont présentés successivement avec leur analyse et interprétation. La synthèse globale est faite dans la partie identification des conceptions qui vient après cette section. Sur l’ensemble des 20 items retenus 14 ont présenté des performances dont les raisons évoquées par les élèves marocains et québécois sont teintées de connotations culturelles. Les 6 autres items présentent des performances dont les raisons ont une connotation purement académique. Par conséquent, deux items ont été choisis parmi les 14 parce que leurs différences de performance et de certitude ont été les plus importantes mais aussi pour montrer l’impact des habitudes de vie quotidienne sur les conceptions qui sont derrière les choix de réponse et qui sont connus d’avance par la communauté scientifique. La suite

présente les résultats obtenus pour chacun des deux items, leur analyse et leur interprétation.

#### **Exemple d'item 4**

La préparation de la pâte à pain est présentée dans un contexte où un cuisinier mélange la farine, l'eau, le sel et la levure. Après le mélange, la pâte est placée dans un conteneur pendant plusieurs heures afin d'accélérer la fermentation. Dans ce contexte, la montée de la pâte à pain est le principal problème à résoudre. Ce problème est présenté comme suit: pendant la fermentation, un changement chimique a lieu dans la pâte: la levure (un champignon unicellulaire) aide à transformer l'amidon et les sucres de la farine en dioxyde de carbone et en alcool. La fermentation fait monter la pâte. La tâche à réaliser consiste à expliquer pourquoi la pâte monte-t-elle? Voici les réponses possibles, la consigne est de souligner une seule réponse.

- A) La pâte monte parce que l'alcool produit se transforme en gaz.
- B) La pâte monte à cause des champignons unicellulaires qui se reproduisent.
- C) La pâte monte parce qu'un gaz, le dioxyde de carbone, est produit.
- D) La pâte monte parce que la fermentation transforme l'eau en vapeur.

La figure ci-après présente les résultats de l'analyse quantitative de l'item 4.

#### **Résultats des entretiens**

Le tableau ci-après récapitule les effectifs des élèves marocains et québécois interviewés en fonction des balises du cadre théorique et des indicateurs de codage de verbatim.

**Tableau 5 : Effectif des répondants par indicateur de codage**

Critère	Indicateur de codage de verbatim	Élèves (Maroc)	Élèves (Québec)	Observation
Concept principal (C)	Étranger au problème	1	3	Sans objet
	Pertinent au problème	17	15	
Représentation (L)	Réfère à ressource médiatique	1	3	Différence importante
	Réfère à une pratique sociale	17	0	
	Réfère au savoir scolaire	0	15	
Problème (P)	Ne reformule pas le problème	2	1	Sans objet
	Reformule le problème	16	17	
Opérateur (R)	Exploite le texte	1	16	Différence importante
	Logique personnelle	17	2	
	Renvoie au texte de la question	0	0	
Validation ( $\Sigma$ )	Exprime un doute	1	15	Différence importante
	Exprime une certitude	17	3	
Réponse	Bonne	1	16	Différence importante
	Mauvaise	17	2	

#### **Au Québec**

Parmi les 18 élèves québécois interviewés, 16 ont proposé des réponses en se basant sur

une exploitation du texte et deux se sont démarqués des autres par leur logique personnelle. Exemple de forme de réponses construites par décodage du texte:

- La levure doit transformée l'amidon, le sucre et la farine en dioxyde de carbone et de l'alcool. (2/16 élèves)
- La pâte monte à cause du gaz qui est créé par la levure. (10/16 élèves)
- Dans le problème, c'est qu'il y a un changement chimique quand je mélange la levure qui est un champignon unicellulaire. (1/16 élève)

Trois (3) répondants parmi les 16 ont expliqué le dégagement du gaz carbonique en se référant à la pâte qui est chauffée dans un four. Exemple de réponse:

- Je pense que ça va se faire comme quand on le met au four. Ben, la réaction chimique, je ne suis pas sûr.
- Parce que quand la pâte est dans le four, il y a une fermentation qui se fait puis il y a une graisse qui se fait puis le dioxyde de carbone se produit.
- J'ai compris qu'on met la pâte à pain dans le four. Ensuite elle se transforme en champignon unicellulaire puis en alcool.

Tous les 16 élèves ayant exploité le texte semblent douter de leur choix de réponse.

Exemple de forme de réponses construites par logique personnelle (2/18 élèves):

- Je trouve que c'est un peu mélangeant je ne suis pas clair avec la question. Parce qu'ici quand je lis pendant la fermentation il y a un changement chimique.
- Moi je pense que c'est parce que quand la pâte est dans le four il y a une fermentation qui se fait puis il y a une graisse qui se fait puis le dioxyde de carbone se produit.

Le doute majoritairement (15 élèves sur 18) exprimé pourrait être issu d'un manque de référent empirique connu de tous les élèves en matière de fabrication du pain. Ces élèves citent souvent le four comme lieu de fermentation ou la chaleur du four comme source de fermentation. Le choix majoritaire (38/63) de la réponse «C» (La pâte monte parce qu'un gaz, le dioxyde de carbone, est produit.) lors de l'écrit semble avoir un lien avec l'idée plus souvent récurrente du gaz carbonique de la fumée qui se dégage dans le four.

### *Au Maroc*

Pour comprendre le système de représentation (L) des élèves marocains, nous nous basons sur le fait que 17 élèves interviewés sur 18 se sont référés à leurs pratiques quotidiennes de fabrication du pain à la maison. Ces 17 élèves soutiennent leurs choix de la réponse « B » avec une certitude élevée. La principale réponse qui émerge de la reformulation du problème est «la levure fait monter la pâte». Cette idée est exprimée majoritairement par 16 élèves sur 18.

Ce résultat confirme les tendances observées lors du test écrit. L'idée qui découle des pratiques quotidiennes de fabrication du pain a été choisie par 39 élèves sur 90 qui avaient passé le test écrit. Parmi les 39 élèves, 31 ont exprimé une certitude élevée. Cependant, l'idée qui découle de l'exploitation du texte, à savoir le choix de la réponse «A» (la pâte monte parce que l'alcool produit se transforme en gaz), n'avait été choisie que par huit élèves lors du test écrit.

Voici quelques exemples de forme de réponses les plus souvent données par

l'échantillon (17/18) des élèves marocains interviewés :

- On peut dire une transformation de la levure. D'abord la pâte n'était pas trop fermentée. Mais après quand on a fait la levure elle s'est fermentée donc c'est une transformation.
- Pour faire du pain on doit toujours faire de la levure chimique pour qu'il transforme la forme de la pâte. (10/17 élèves)
- La pâte monte parce qu'il contient de la levure chimique. (2/17 élèves)
- La pâte monte parce qu'il y a la levure. (5/17 élèves)

L'idée récurrente de la levure chez les élèves marocains dans le processus de la fabrication de la pâte à pain suggère qu'il semble exister une prédominance des connaissances culturelles sur les connaissances scolaires dans la réalisation de la tâche demandée.

### ***Interprétation didactique***

Les élèves marocains (17/18) et québécois (16/18) affirment avoir étudié la fermentation en classe. Ces élèves avaient donc, du moins en théorie, un même potentiel académique dans la réalisation de la tâche demandée. Les élèves marocains soutiennent, en plus, qu'ils connaissent les ingrédients et le processus de la fabrication de la pâte à pain. Pour plusieurs élèves marocains (17/18), on trouve deux idées simultanément: (1) l'idée qui découle des pratiques quotidiennes de fabrication du pain (L): la pâte monte à cause de la levure et (2) l'idée qui découle de l'exploitation du texte notamment le choix de la réponse (R): la pâte monte parce que l'alcool produit se transforme en gaz. Ces élèves se sont donc référés à leur pratique quotidienne de fabrication du pain pour réaliser leur tâche.

En conséquence, la formulation du contexte dans cet item pourrait être déterminante dans l'explication de la différence de performance constatée entre les élèves marocains et québécois. Les résultats indiquent que pour cet item 4, le lien entre la reformulation du problème (P) et la logique du choix de réponse choisie (R) est incohérent. Cette incohérence se révèle dans le lien entre la reformulation du problème, autrement dit l'explicitation par l'élève de sa compréhension de la tâche à réaliser et son choix de réponse.

Ces résultats présentent également un caractère paradoxal en ce qu'ils permettent de voir que dans le cas de l'item 4, s'éloignent de la bonne réponse les élèves qui sont en réalité les plus proches du phénomène étudié (c'est-à-dire : la fabrication du pain) ou alors les élèves québécois par un manque de familiarité avec le phénomène se rabattent sur l'exploitation de texte et produisent la bonne réponse.

### ***Exemple d'item 14***

Cet item ne présente pas un contexte détaillé du problème. La tâche demandée se présente comme suit : Comment la température moyenne du corps des êtres vivants des climats chauds se compare-t-elle à la température moyenne du corps des êtres vivants des climats froids? Voici les réponses possibles, la consigne consiste à souligner une seule réponse.

A) Plus haute dans les climats chauds; B) Plus basse dans les climats chauds; C) La même dans les deux climats

### ***Résultats des entretiens***

Les données du tableau ci-après indiquent que les différences en termes d'effectifs de répondants n'ont pas été suffisamment importantes pour en tirer d'emblée des conclusions

pertinentes. L’item 14 avait été retenu parce que les résultats à l’écrit révélaiement une différence de choix de réponse. Les élèves marocains (44/89) ont majoritairement choisi la réponse «A» tandis que les élèves québécois (45/89) ont majoritairement choisi la réponse «C». Ces tendances ne se sont cependant pas observées au niveau des propos recueillis lors des entretiens.

**Tableau 6 : Effectif des répondants par indicateur de codage**

Critère	Indicateur de codage de verbatim	Élèves (Maroc)	Élèves (Québec)	Observation
Concept principal (C)	Étranger au problème	0	0	Sans objet
	Pertinent au problème	18	18	
Représentation (L)	Réfère à ressource médiatique	0	1	Sans objet
	Réfère à une pratique sociale	16	15	
	Réfère au savoir scolaire	2	2	
Problème (P)	Ne reformule pas le problème	4	1	Sans objet
	Reformule le problème	14	17	
Opérateur (R)	Exploite le texte	0	0	Sans objet
	Logique personnelle	18	16	
	Renvoie au texte de la question	0	2	
Validation ( $\Sigma$ )	Exprime un doute	7	8	Sans objet
	Exprime une certitude	11	10	
Réponse	Bonne	9	6	Sans objet
	Mauvaise	9	12	

Quelques exemples de forme de propos tenus par les dix-huit (18) élèves marocains:

- La température moyenne des êtres vivants les climats chauds. Je pense que la température est chaude. C’est pourquoi le corps des êtres vivants devient très chaud. (2/18 élèves)
- La température du corps se transforme avec la température moyenne lorsqu’il fait chaud. Si la température du corps va se transformer. Lorsqu’il fait chaud un petit peu la température du corps augmente. (12/18 élèves)
- Il y a la chaleur pour la température moyenne du corps de l’être vivant le climat chaud. Pour la température moyenne des êtres vivants les climats froids, il n’y a pas assez de chaleur. (4/18 élèves)

Ces propos témoignent que les élèves marocains s’inspirent de leur contexte géographique de vie pour justifier leur choix de réponse.

Quelques exemples de forme de propos tenus par les élèves québécois (16/18):

- C’est la même dans les deux climats parce que le corps a pas mal une seule température peu importe où on vit ça se peut c’est juste notre peau qui est capable de supporter la température ou pas mais en dedans de nous ça sera toujours la même température. (10/18 élèves)
- Par exemple de l’été à l’hiver, comment notre corps réagit-elle à cette situation-là?

Je dirai que c'est la même dans les deux climats mais ça peut aussi dépendre des gens. Les gens par exemple, à l'été, ils aiment plus ça parce qu'il fait plus chaud et que le corps est déjà habitué à la chaleur sur eux. Mais parfois à l'hiver, ils sont moins habitués à avoir le froid sur eux donc le corps ne réagit pas à la même réalité. La température du corps va augmenter et le corps va devenir malade. Généralement, comme une personne comme moi, on s'habitue dans les deux climats. Par exemple, comme si nous sommes en été, en ce moment il fait tellement chaud ton corps est déjà habitué. Et aussi le pays ça dépend du climat du pays. Si le climat est resté toujours chaud, ton corps va peut-être l'accepté mais si le climat est chaud puis froid ton corps a besoin de ça pour être vraiment ok. (4/18 élèves).

- Ça, je ne connais pas. Ben, on m'a dit que la température moyenne d'un corps c'est 37°C. Mais, je ne sais pas si dans un climat froid si elle descend ou elle monte. Je sais que la température du corps est supposée rester autour de 37°C mais je ne sais pas si elle diminue où elle monte. (2/18 élèves)

Ces propos témoignent que les élèves québécois se sont également inspirés de leur contexte géographique de vie pour justifier leur choix de réponse. L'alternance entre l'été et l'hiver permet possiblement aux élèves de tenter de comprendre leur adaptation climatique et de répondre conséquemment à la question posée.

### ***Interprétation didactique***

Une remarque préalable dans l'énoncé de l'item, le terme «êtres vivants» englobe plusieurs espèces animales comme végétales. Son usage pour préciser la constance de la «température moyenne interne» n'est pas généralisable. Cette température moyenne varie naturellement en fonction des êtres et de leur milieu de vie.

Les élèves marocains déclarent avoir l'impression que la chaleur ressentie sur la peau serait la même que dans le corps. Cette représentation (L) de la température interne est celle des Ectothermes. Ces derniers sont des organismes dont la température interne est identique à celle du milieu ambiant. C'est le cas des insectes, des reptiles et des poissons qui sont massivement présents dans leur milieu de vie. Par conséquent, le choix de la réponse «la température moyenne est plus haute dans le climat chaud» serait la bonne réponse dans leur contexte géographique de vie.

Dans le contexte géographique de vie des élèves québécois, il n'y a presque pas d'animaux vertébrés terrestres à sang froid. Les animaux vertébrés terrestres qu'on y trouve sont prototypiques pour ces élèves et ceux-ci s'adaptent aux changements de climat chaud et froid. Il serait donc compréhensible que dans leur contexte géographique de vie, les élèves québécois affirment que «la température moyenne interne des êtres vivants soit la même».

Par conséquent, le contexte et les distracteurs proposés dans cet item peuvent induire des biais qui sont aussi déterminant au même titre que les différences de contextes culturels dans l'explication des différences de performance à cet item. Pour saisir la scientificité de la tâche demandée, un potentiel académique est nécessaire car la seule expérience de vie ne semble pas être suffisante pour trouver la bonne réponse.

De même, les distracteurs tels que : «plus haut dans les climats chauds» ou «plus basse dans les climats froids» semblent avoir un lien de cause à effet avec le caractère relatif de la température interne du corps humain. C'est ce lien possible qui pourrait être à l'origine des résultats des élèves de deux groupes.

Le contexte géographique de vie s'est donc avéré une ressource importante dans la réalisation de la tâche. Or dans la construction des items standardisés et décontextualisés utilisés dans les évaluations internationales, les organisateurs postulent qu'aucune influence contextuelle ou culturelle ne devrait orienter les réponses des élèves. Le cas échéant, l'estimation du maximum de vraisemblance permettrait de minimiser ces biais culturels. L'exemple des résultats de cet item contraste donc ce postulat.

### **Identification des conceptions**

Cette section réexamine de façon minutieuse les liens entre le problème présenté dans l'item, la réponse majoritairement choisie par un groupe et l'argument majoritairement évoqué pour soutenir ce choix (Harlow & Jones, 2004). La conception sera dite avérée lorsque la cohérence entre les trois éléments P, R, L est vérifiée. Cette conception avérée sera qualifiée de rigide ou de fragile selon le degré de certitude majoritairement exprimée. Une discussion sera engagée pour établir les liens possibles entre une conception identifiée et les distracteurs proposés dans les choix de réponse.

### ***L'item à propos de la préparation de la pâte***

Cet item avait été retenu à cause de la divergence de choix de réponse entre les deux groupes. Le problème (P) présenté dans cet item est: Pourquoi la pâte à pain monte-t-elle?

### ***Au Québec***

Pour l'opérateur (R), le choix majoritaire à l'écrit (38/63 élèves) est la réponse «C» (La pâte monte parce qu'un gaz, le dioxyde de carbone, est produit). Ce choix a été soutenu lors des entretiens par (16/18) élèves. Ces élèves affirment qu'ils ont procédé par une exploitation de texte. (10/18) élèves justifient leur choix lors des entretiens par la réponse: La pâte monte à cause du gaz qui est créé par la levure. (3/18) justifient leur choix par la réponse: La pâte à pain monte à cause du dioxyde de carbone produit dans le four. Ces éléments de justification contenus dans leur système de représentation (L) semblent refléter une intuition basée sur le décodage du texte avec quelques appuis culturellement teintés. (15/18) de ces élèves n'étaient pas sûrs de leur choix de réponse.

Ces résultats confirment les tendances observées à l'écrit. La cohérence entre les balises P, R, L n'est pas confirmée par le système de validation ( $\Sigma$ ). En conséquence, l'idée du gaz qui fait monter la pâte à pain n'est pas une conception avérée car les élèves doutent de leur choix de réponse.

### ***Au Maroc***

Pour l'opérateur (R), la réponse majoritairement (39/90 élèves) choisie est «B» (La pâte monte à cause des champignons unicellulaires qui se reproduisent). Ce choix a été soutenu lors des entretiens par (17/18) élèves. Ces élèves affirment se référer à leur pratique quotidienne de fabrication du pain. Ils soutiennent leur choix de réponse lors des entretiens à partir des réponses suivantes:

- Pour faire du pain, on doit toujours mettre de la levure chimique pour qu'il transforme la forme de la pâte. (10/17 élèves)
- La pâte monte parce qu'il contient de la levure chimique. (2/17 élèves)
- La pâte monte parce qu'il y a la levure. (5/17 élèves)

Ces éléments de leur système de représentation (L) traduisent une connaissance valide dans

leur contexte culturel. Ces élèves (17/18) soutiennent leurs propos avec une certitude élevée.

Ces résultats confirment les tendances observées à l'écrit. En conséquence, l'idée de la levure fait monter la pâte à pain serait ancrée dans le contexte culturel marocain car le système de validation  $\Sigma$  confirme la certitude élevée.

Pour identifier la conception qui découle des propos des élèves marocains, nous allons procéder à la juxtaposition des caractéristiques P, R, L et  $\Sigma$  afin d'établir si possible la cohérence. Le lien entre le problème (P) (pourquoi la pâte à pain monte-t-elle ?), la réponse majoritairement choisie à l'écrit (R) (La pâte monte à cause des champignons unicellulaires qui se reproduisent), l'argument ayant soutenu ce choix (L) (Pour faire du pain on doit toujours mettre la levure chimique pour qu'il transforme la forme de la pâte), et le système de validation ( $\Sigma$ ) est cohérent et direct. Les caractéristiques P, R, L et  $\Sigma$ , dans ce cas, sont vérifiées.

Dans le contexte culturel marocain, la pâte à pain monte parce que des champignons unicellulaires se reproduisent. Or la levure est faite des champignons unicellulaires. L'élève marocain soutient avec une certitude élevée que pour faire du pain on doit toujours mettre la levure chimique pour qu'elle transforme la pâte.

### *L'item à propos de la température moyenne*

L'item avait été retenu à cause de la divergence de choix de réponse entre les deux groupes. Les élèves marocains (44/89) ont majoritairement choisi la réponse «A» (La température moyenne des êtres vivants est plus haute dans les climats chauds) tandis que les élèves québécois (45/89) ont majoritairement choisi la réponse «C» (La température moyenne des êtres vivants est la même dans les climats chaud et froid).

Le problème (P) présenté dans cet item est: comment la température moyenne du corps des êtres vivants des climats chauds se compare-t-elle à la température moyenne du corps des êtres vivants des climats froids ?

Les éléments de représentation majoritairement soutenus lors des entretiens sont:

- Pour le Québec: C'est la même dans les deux climats parce que le corps a pas mal une seule température peu importe où on vit. ça se peut. C'est juste notre peau qui est capable de supporter la température ou pas mais en dedans de nous ça sera toujours la même température. (10/18 élèves)

Pour le Maroc

- La température du corps se transforme avec la température moyenne lorsqu'il fait chaud. Lorsqu'il fait chaud, la température du corps augmente un petit peu. (12/18 élèves)

Ces propos sont soutenus avec une certitude élevée. Ces données confirment les résultats obtenus lors de l'écrit. Pour identifier des conceptions qui en découlent, nous allons juxtaposer les caractéristiques P, R, L et  $\Sigma$  recueillies pour chaque groupe.

Au Québec, les élèves affirment avec une certitude élevée que la température moyenne des vivants les climats chauds est la même que pour les vivants les climats froids. Ils soutiennent leurs propos en se référant à leur propre adaptation au climat chaud (l'été) et au climat froid (l'hiver).

Au Maroc, les élèves affirment avec une certitude élevée que la température moyenne des êtres vivants les climats chauds est plus haute que celle des êtres vivants les climats froids. Ils soutiennent leurs propos en se référant à leur propre mode de vie. Ces

deux conceptions sont contextuellement valides. Le caractère contextuel des conceptions des élèves sur la notion de température avait été déjà prouvé par Colin et Tran Tat (2011), sur l'étude des conceptions d'élèves à propos de l'effet de serre. À l'instar des travaux de Laval (1985) sur les conceptions de température, ce résultat donne un important élément de réponse pertinente à notre question globale de recherche.

En conséquence, une leçon à tirer de cette discussion est que l'ensemble des conceptions en sciences contextuellement valides identifiées au cours de cette étude qualitative de différences de performance à une évaluation internationale permettent de soutenir affirmativement la réponse à notre question globale de recherche. Les différences de performance par item constatées ci-dessus (voir tableau 1) ne relève pas vraiment du hasard notre étude y apporte quelques pistes d'explication.

À l'instar de Chang, Yeh et Barufaldi (2010), Othman, Treagust et Chandrasegaran (2008), nous disons oui, il existe des conceptions en sciences pouvant être à l'origine des réponses d'élèves et qui, d'après notre étude, sont susceptibles d'expliquer les différences de performance qu'on observe lors de la réalisation d'une tâche d'évaluation internationale. Conséquemment, l'estimation de degré de certitude aux réponses a été un outil qui a permis de comprendre que le caractère contextuel des connaissances construites au sein des communautés de pratique de n'importe quel contexte culturel ou système scolaire (Chu, Treagust, Yeo, & Zadnik, 2012) est tributaire de certaines différences de performance observées dans la réalisation des tâches standardisées.

## **Conclusion**

Au regard des limites de l'échantillon et du nombre d'items utilisés pour construire une interprétation plus soutenue, nos résultats ne peuvent être généralisés à l'ensemble des enquêtes internationales. Cependant, il ressort de notre étude que le contexte culturel influe sur le rendement des élèves dans la réalisation des tâches décontextualisées.

Ce postulat didactique auquel nous sommes parvenus permet d'amorcer une perspective qualitative d'explication des résultats enquêtes internationales. Cette assertion est soutenue par la catégorie des réponses obtenues. La première catégorie de réponses indique la persistance de l'influence du contexte culturel dans les réponses d'élèves malgré que les tâches soient standardisées. La deuxième catégorie des réponses indiquent l'existence d'un potentiel académique qu'offre un système scolaire à ses élèves pour que ceux-ci soient capables de lire et de comprendre le vocabulaire du contenu des items décontextualisés.

Cette catégorisation permet de faire émerger deux vecteurs qui orientent la formulation des hypothèses didactiques sur les résultats obtenus. Ces deux vecteurs sont: (1) l'influence du contexte culturel; (2) le potentiel académique.

L'exemple de la fabrication de pâte à pain dans l'item 4 est éloquent : des élèves capables de faire cuire du pain (Marocains) vont moins bien performer aux tests internationaux que des élèves qui en sont incapables (Québécois). Ces exemples de propos attestent que des efforts supplémentaires devraient être consentis pour assurer l'équité dans les tâches internationales.

Une piste probable d'équité dans la réalisation des tâches internationales serait l'adaptation des items aux contextes culturels. Possiblement, la prise en compte des contextes culturels nationaux dans l'analyse des items pourrait résulter en une relativisation de la valeur comparative des tests internationaux. Ces derniers prétendent évaluer

l'efficacité des systèmes scolaires, mais se heurtent aux conceptions issues des communautés de pratiques d'où émergent les réponses.

Finalement, notre étude se termine en esquissant quelques hypothèses d'ordre général pouvant ouvrir des recherches sur la problématique de l'équité aux évaluations internationales:

(1) Lorsque les contextes et les moyens de construction des connaissances établissent des différences entre des groupes d'élèves, la distanciation observée au sein des communautés de pratique pourrait être tributaire des différences de performance.

(2) Dans la construction et la sélection des items aux évaluations internationales, l'estimation de degré de certitude aux réponses peut fournir au préalable des données qualitatives susceptibles de compléter et d'enrichir par la suite les données contextuelles recueillies auprès des administrations, les élèves et les parents de chaque pays afin de dégager la valeur qualitative des comparaisons.

### Références

- Allal, L. (2007). Évaluation dans le contexte de l'apprentissage situé : peut-on concevoir l'évaluation comme un acte de participation à une communauté de pratiques ? Dans M. Behrens, *La qualité en éducation : pour réfléchir à la formation de demain* (pp. 39-56). Québec: Presses de l'Université du Québec.
- Angell, C. (2004). Exploring Students' Intuitive Ideas Based on Physics Items in TIMSS 1995. In C. Papanastasiou (éd.), *Proceedings of the IRC-2004 TIMSS* (pp. 108-123). Nicosia: Cyprus University Press.
- Angell, C., Kjærnsli, M., & Lie S. (2006). Curricular and cultural effects in patterns of students' responses to TIMSS science items. In S. J. Howie & T. Plomp (éd.), *Contexts of learning mathematics and science: Lessons learned from TIMSS* (pp. 277-290). London: Routledge.
- Balacheff, N. (1995). Conception, propriété du système sujet/milieu. In R. Noirfalise., M.-J. Perrin-Glorian (eds.) *Actes de la VII École d'été de didactique des mathématiques* (pp. 215-229). Clermont-Ferrand: IREM de Clermont-Ferrand.
- Boutin, G. (2000). *L'entretien de recherche qualitatif*. Sainte-Foy Presses de l'Université du Québec.
- Bulles, N (2010, mars-avril). L'imaginaire réformateur. PISA et les politiques de l'École. *Le Débat*, 159, 95-109.
- Caleon, I., & Subramaniam, R. (2010). Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions. *Research in Science Education*, 40(3), 313-337.
- Chang, C. Y., Yeh, T. K., & Barufaldi, J. P. (2010). The positive and negative effects of science concept tests on student conceptual understanding. *International Journal of Science Education*, 32(2), 265-282.
- Chen, J.-Y., Chang, H.-P., & Guo, Ch.-J. (2004). The Development of a Diagnostic Instrument to Investigate Students' Alternative Conceptions of Reflection and Refraction of Light (Article written in Chinese). *Chinese Journal of Science Education*, 12(3), 311-340
- Chu, H.-E., Treagust, D. F., Yeo, S., & Zadnik, M. (2012). Evaluation of Students' Understanding of Thermal Concepts in Everyday Contexts, *International Journal of Science Education*, 34(10), 1509-1534
- Colin, Ph., & Tran Tat, N. (2011). Difficile compréhension de l'effet de serre : comment concevoir un parcours d'enseignement-apprentissage au lycée?, *RDST*, 4, 109-138
- Foy, P., Arora, A., & Stanco, G.M. (Eds.). (2013). *TIMSS 2011 user guide for the international database*. Chestnut Hill, MA: Boston College.
- Gilles, J.-L. (2002). Qualité spectrale des tests standardisés universitaires. Thèse de doctorat inédite en Sciences de l'éducation. Liège : Université de Liège
- Giordan, A., & De Vecchi, G. (1987). Les origines du savoir. Des conceptions des apprenants aux concepts scientifiques, Delachaux et Niestlé
- Goldstein, H. (2004). International comparative assessment: how far have we really come? *Assessment in Education*, 11(2), 227-234.
- Grønmo, L. S., Kjærnsli, M., & Lie, S. (2004). Looking for cultural and geographical factors in patterns of response to TIMSS items. In C. Papanastasiou (éd.), *Proceedings of the IRC-2004 TIMSS*. Nicosia: Cyprus University Press, 1, 99-112.

- Gonzalez, E. J. (2003). Scaling the PIRLS reading assessment data. In I. V. S. Mullis, M. O. Martin, E. Gonzalez & A. Kennedy, *PIRLS 2001 International Report*. Boston: International Study Center.
- Grisay, A., Gonzalez, E., & Monseur, C. (2009). Equivalence of Item Difficulties Across National Versions of the PIRLS and PISA Reading Assessments. *IERI Monograph Series, 2*, 63-83.
- Harlow, A., & Jones, A. (2004). « Why Students Answer TIMSS Science Test Items the Way They Do ». *Research in Science Education, 34*(2), 221-238.
- Kaplan, D. (2004). *The SAGE Handbook of Quantitative Methodology for the Social Sciences*. University of Wisconsin – Madison: SAGE Publications, Inc.
- Kjærnsli, M., Angell, C., & Lie, S. (2002). « Exploring Population 2 Students' Ideas about Science ». In D. F. Robitaille & A. E. Beaton (éd.), *Secondary Analysis of the TIMSS Data*. Dordrecht: Kluwer, p. 127-144.
- Kraus, P., & Minstrell, J. (2002). Designing Diagnostic Assessments. *Proceedings of the Physics Education Research Conference*. Boise, ID.
- Laval, A. (1985). Chaleur, température, changements d'état. In *Chaud, froid, ... pas si simple : quelques activités concernant les notions de température et de chaleur* (pp.115-132). Paris: INRP
- Martin, M.O., & Mullis, I.V.S. (Eds.). (2012). *Methods and procedures in TIMSS and PIRLS 2011*. Chestnut Hill, MA: Boston College.
- Martin, M.O., Mullis, I.V.S., Foy, P., & Stanco, G.M. (2012). *TIMSS 2011 international results in science*. Chestnut Hill, MA: Boston College.
- Mons, N. (2008). Evaluation des politiques éducatives et comparaisons internationales: introduction. *Revue française de pédagogie, 164*, 5-13.
- OCDE (2010). *Les Clés de la Réussite : Impact des connaissances et compétences à l'âge de 15 ans sur le parcours des jeunes Canadiens*. Paris: Éditions OCDE.
- Orange, C., & Orange Ravachol, D. (2013). Le concept de représentation en didactique des sciences : sa nécessaire composante épistémologique et ses conséquences, *Recherches en éducation, 17*, 46-61.
- Othman, J., Treagust, D.F., & Chandrasegaran, A. L. (2008): An Investigation into the Relationship between Students' Conceptions of the Particulate Nature of Matter and their Understanding of Chemical Bonding, *International Journal of Science Education, 30*(11), 1531-1550.
- Park, C., & Bolt, D. M. (2008). Application of multi-level IRT to investigate cross-national skill profiles on TIMSS 2003. *IERI monograph series: Issues and methodologies in large-scale assessments, 1*, 71-96.
- Tsai, C.-C., & Chou C. (2002). Diagnosing students' alternative conceptions in science. *Journal of Computer Assisted Learning, 18*, 157-165.
- Turmo, A. (2003, Août). Understanding a newsletter article on ozone: a cross-national comparison of the scientific literacy of 15-year-olds in a specific context. Communication présentée à la 4e conférence ESERA Research and the Quality of Science Éducation, Noordwijkerhout [Pays-Bas].
- Simard, C., Harvey, L., & Samson, G. (2014). Regard multidimensionnel des conceptions du vivant : situation en contexte québécois, *RDST, 9*, 79-102.
- Wu, M. L. (2008, September). A Comparison of PISA and TIMSS 2003 achievement results in Mathematics and Science. Paper presented at the Third IEA Research Conference, Taipei.

**Guy Norbert LOUBAKI, Ph.D**, est chercheur à l'Institut National de Recherche et d'Action Pédagogique (INRAP-Congo) professeur de mathématiques et sciences au secondaire à la Commission scolaire de Kativik au Québec, consultant international en éducation scientifique et technologique, membre de l'Équipe de recherche en éducation scientifique et technologique (EREST) à l'université du Québec à Montréal.

**Patrice Potvin, Ph.D**, est professeur en didactique des sciences et de la technologie au secondaire à la Faculté des sciences de l'éducation de l'Université du Québec à Montréal. Ancien enseignant au primaire et au secondaire, il est titulaire de la Chaire de recherche sur l'intérêt des jeunes à l'égard des sciences et de la technologie (CRIJEST), directeur du Laboratoire Mobile pour l'étude des cheminements d'apprentissage en science (LabMECAS) et de l'Équipe de recherche en éducation scientifique et technologique (EREST).

**Lalla Rabia HIJAZI, Ph.D**, est Professeure de didactique des mathématiques. À l'École Normale Supérieure Takaddoum à Rabat, Maroc

**Jesús Vázquez-Abad, Ph.D**, est professeur agrégé, titulaire de la Chaire d'étude et de recherche en enseignement des sciences et technologies en milieu scolaire et collégial à l'Université de Montréal, coresponsable du programme de la maîtrise en enseignement au secondaire