

Regard didactique sur l'explication des différences de performance des élèves marocains et québécois aux évaluations internationales

Guy Norbert LOUBAKI

Université du Québec à Montréal, loubaki2@yahoo.fr

Patrice POTVIN

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

Vàzquèz-Abad JESUS

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

Résumé

Cet article présente une étude en didactique des sciences qui a eu pour objectif de mieux comprendre et expliquer les différences de performance réalisées par des jeunes marocains et québécois de 15 ans dans une évaluation internationale. Notre étude s'inscrit dans une perspective de diagnostic des conceptions en sciences (Tsai et Chou, 2002, Kraus et Minstrell, 2002 ; Thijs et Van Den Berg, 1995 ; Tsai et Chou, 2002). Le cadre théorique est inspiré des travaux de Balacheff (1995) sur la caractérisation des conceptions suivant la dynamique sujet/milieu. La démarche de recherche est basée sur une approche mixte à dominance qualitative. Nos résultats attestent l'existence d'une importante influence de conceptions d'origine culturelle dans l'orientation des réponses d'élèves aux items standardisés utilisés dans certaines grandes enquêtes internationales. Ces résultats offrent donc des pistes de réflexion sur la possibilité de relativiser les explications de différence de performance entre pays de contexte culturel différent ainsi que des pistes de solution pour les minimiser. Conséquemment, notre étude soulève un questionnement de nature didactique sur la modélisation des contenus et de format d'items dits standardisés. Quels peuvent être les invariants dans l'habillage d'un item standardisé en sciences au regard des conceptions contextuellement valides qui orientent les réponses des élèves?

Mots clés : conceptions en sciences, évaluation internationale, performance

I Contexte d'étude

Cette étude se réalise dans un contexte où les performances des pays aux évaluations internationales en éducation (PISA, TEIMS, Sacmeq, Pasec...) sont sujettes aux diverses interprétations occasionnant parfois des vives critiques de formats ou de contenus d'items proposés. Les interprétations divergent selon le rôle et la place attribués au contexte culturel qui conditionnent la construction de connaissances et les réponses données par l'élève. Les problèmes inhérents à la manière dont les items des tests internationaux sont construits, notamment leur format et leur contenu, motivent cette étude didactique. Plus particulièrement, cette étude s'intéresse aux difficultés qu'éprouvent les élèves de contextes géographiques différents à pouvoir réaliser de façon équitable des tâches demandées.

Dans la littérature portant sur les grandes enquêtes internationales, les différences de score de performance souvent récurrentes à travers diverses sessions, tel que l'indique le tableau ci-dessous, n'ont pas toujours trouvé d'explication du point de vue didactique.

Contexte géographique	Pays	2003	2007
Nord-américain (groupe B)	USA	527	520
	Québec	531	507
Asiatique (groupe A)	Japon	552	554
	Singapour	578	567
Maghrébin (groupe C)	Maroc	396	402
	Tunisie	404	445

Tableau 1 : Performances extraites des résultats TEIMS 2003 et 2007 sciences (NCES, 2008)

Ces données indiquent de façon récurrente que les élèves des pays asiatiques performant toujours mieux aux évaluations internationales que les élèves des pays nord-américains ou européens. Il est donc curieux de constater que tous les pays arabes, par exemple, soient souvent moins performants aux évaluations internationales de type PISA et TEIMS alors que les pays asiatiques sont toujours meilleurs. Ces variations sont souvent maintenues mêmes au sein des pays multiculturels. A ce sujet, Clarke (2006), se basant sur une étude des performances des élèves américains issus de familles immigrées de pays asiatiques, fait constater qu'aux États-Unis, ces élèves sont aussi performants que ceux d'Asie. Ce démarquage de performance laisse penser que les élèves des pays asiatiques, des pays arabes et des pays nord-américains voire européens ont certainement diverses manières de conceptualiser les phénomènes scientifiques selon leurs habitudes de vie quotidienne.

Dans cet ordre d'idée, Wu (2008) fait constater que pour un même contenu notionnel et une mise en situation différente entre les items de TEIMS 2003 et de PISA 2003 en mathématiques, les élèves des pays occidentaux sont plus performant sur les tests PISA que les élèves des pays asiatiques tandis que les élèves des pays asiatiques sont plus performant sur les tests TEIMS que les élèves des pays occidentaux (Wu, 2008). Les conclusions de Wu renvoient aux constats faits par Park & Bolt (2008), sur la manière dont les items de tests internationaux sont construits (choix des tâches et format des items).

Ce constat fait par Wu (2008), sur les performances des élèves asiatiques et européens, laisse penser que malgré que les tâches d'évaluation internationales soient décontextualisées, il existerait des conceptions en sciences dépendantes des contextes géographiques de vie qui pourraient orienter les réponses des élèves lors de la réalisation de ces tâches.

Allal (2007), fait également constater 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 donc des difficultés à établir l'équité sinon à expliquer les différences de performances observées. En conséquence, l'auteure soutient que les élèves de contextes différents ont nécessairement acquis leurs savoirs et compétences grâce à leur insertion dans des communautés de pratiques (Lave, 1988) extrêmement variables (Allal, 2007).

Par conséquent, ces constats permettent de circonscrire dans la suite un problème de recherche en didactique des sciences des explications des différences de performance des élèves de contextes géographiques différents à partir des conceptions de toute nature susceptibles d'orienter leurs réponses aux items des évaluations internationales.

2 Problème de recherche

Pour cerner le problème de recherche en didactique dont il est question dans cette étude, nous présentons d'abord les raisons soutenues par des spécialistes en éducatrice qui mettent en exergue les insuffisances des valeurs explicatives des approches quantitatives et qui prônent la prise en compte des approches qualitatives pour affiner puis compléter les explications des performances de pays aux tests internationaux. Des études en éducatrice (Blum, Goldstein & Guerin-Pace, 2001 et Gustafsson & Rosén, 2004) ont donc fait constater que la prise en compte du score de performance comme unique donnée (exigence d'unidimensionnalité) pour inférer les explications d'une variable latente telle que la compétence générale (dans le cas des enquêtes PISA) ne suffit pas car d'autres aspects qualitatifs peuvent contribuer au développement d'une compétence.

Les partisans de cette thèse dont Goldstein (2004), fondent leur critique en se basant sur un exemple d'application de la méthode des équations structurelles aux données françaises et anglaises du PISA 2003. L'auteur affirme à l'issue de son étude que l'écart à l'unidimensionnalité de la variable latente (donnée inobservable : aptitude, compétence, etc.)

ou la référence unique au score de performance pour expliquer la compétence générale (Vrignaud, 2006) est révélateur de limites des modèles explicatifs ayant recours aux valeurs plausibles. Goldstein soutient donc que ces limites risquent d'éliminer des informations porteuses de différences qualitatives supportant d'autres dimensions en complément des différences quantitatives consistant à ordonner les moyennes des pays sur un axe. L'auteur s'interroge ainsi sur la pertinence d'expliquer les différences entre sujets de manière uniquement quantitative.

Cette problématisation des approches quantitatives faite par Goldstein offre des pistes d'approfondissement des réflexions sur l'équité et les valeurs qualitatives des résultats des enquêtes internationales en éducation. Pour transposer ce problème de la généralisabilité des résultats des enquêtes internationales sur le plan didactique, la présente étude passe en revue les perspectives de recherches didactiques des quatre dernières décennies à propos de l'influence du contexte et des conceptions des élèves sur leurs réponses afin de construire le sens didactique de son problème de recherche.

Des travaux en didactiques des sciences (Park & Bolt, 2008; Angell, 2004 ; Dossey, Jones & Martin, 2002 et Kjærnsli, Angell & Lie 2002) ont été axés sur des analyses secondaires de résultats des évaluations internationales. Les auteurs de ces travaux ont tenté par diverses approches quantitatives d'expliquer si les conceptions véhiculées à travers les contenus de réponses écrites par les élèves pouvaient justifier les difficultés rencontrées dans la réalisation des tâches d'évaluation. Leurs conclusions étaient affirmatives.

D'autres études didactiques ont déjà confirmé et soutenu que le contexte de vie quotidienne de l'élève exerce une influence sur ses réponses (Viennot, 1977 et 1989 ; Halloun & Hestenes, 1985 ; Finegold & Gorsky, 1991 ; Thijs & Van Den Berg, 1995 ; Palmer, 1997 et 2001 ; Tao & Gunstone, 1999 ; Kraus & Minstrell, 2002). Certaines études didactiques ont fait également constater la résistance des conceptions dans les réponses données par des élèves à la suite d'un apprentissage (Tsai & Chou, 2002 ; Givry, 2003 ; Treagust & Duit, 2008, Potvin, Mercier, Charland & Riopel, 2011).

En conséquence, un postulat émerge de l'ensemble de ces travaux didactiques. Tous ces auteurs soutiennent le postulat selon lequel le contexte de construction des connaissances d'un élève donne du sens à ses réponses. Ce postulat didactique laisse penser que toute évaluation de connaissances ou de compétences, dans un souci d'équité, devrait prendre en compte les circonstances contextuelles dans lesquelles les connaissances sont construites ou celles dans lesquelles les compétences se développent.

Ce postulat didactique s'oppose donc au fondement même des évaluations internationales qui prônent des contenus d'items décontextualisés pour éviter des biais linguistiques, contextuels ou culturels. Vraisemblablement, il y a donc un contraste entre le postulat didactique sur l'influence du contexte d'émergence des connaissances sur les réponses et les considérations éducatives de l'évitement des biais contextuels sur les réponses des élèves aux évaluations internationales (Lafontaine, 2010). Ce contraste présume immédiatement l'existence d'un problème de recherche en didactique d'explication de performance selon la prise en compte ou non du sens contextuel des réponses.

D'autres études soutiennent la prise en compte de l'estimation du degré de certitude (Caleon et Subramaniam, 2010 ; Gille, 2002 ; Hasan, Bagayoko et Kelley, 1990), comme donnée capitale dans le diagnostic d'une conception et, donc, dans l'explication des différences de performance observées. Ces études ont établi des liens entre le degré de certitude aux réponses et les conceptions qui motivent ces réponses. Cette perspective d'étude offre des nouvelles données encore très peu exploitées par les recherches didactiques au sujet des

évaluations internationales. Ces nouvelles données permettent donc d'affiner la circonscription de notre problème de recherche.

Finalement, notre problème de recherche en didactique des sciences est donc de comprendre et d'expliquer qualitativement les différences de performances aux évaluations internationales à partir des conceptions en sciences des élèves de contexte géographique de vie considérablement différents. Par conséquent, les conceptions en sciences des élèves de contextes géographiques différents dans la réalisation des tâches d'évaluation internationale et leur degré de certitude aux réponses deviennent des objets de recherche susceptibles de nourrir les explications de différence de performance observées. De ce problème de recherche découle des questionnements. Il y a lieu donc de les préciser et de clarifier les objectifs visés.

3 Questions et objectifs de recherche

Du problème de recherche ci-dessus découle la question globale suivante : 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 nécessaires :

- (1) Identifier les 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) Catégoriser les performances des élèves en fonction de leur degré de certitude à la réponse.
- (3) 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 opérationnaliser ces objectifs de recherche, des choix théoriques sont nécessaires afin de trouver un cadre théorique approprié pour analyser les résultats.

4 Choix théoriques

4.1 Cognition distribuée

La théorie de la cognition distribuée postule que :

(...), human cognition is not just influenced by culture and society, but that is in a very fundamental sense a cultural and social process. (...), if groups can have cognitive properties that are significantly different from those of the individuals in them, then differences in the cognitive accomplishments of any two groups might depend entirely on differences in the social organisation of distributed cognition and not to all differences in the cognitive properties of individuals in the groups. (Hutchins, 1995, p. 177-178).

Ce postulat permet le passage d'une vue de la cognition comme un processus/système interne (mental) mis en œuvre par un individu, à une vue de la cognition comme un processus/système distribué, à la fois entre un individu et son environnement physique, et entre cet individu et d'autres avec qui il interagit. De sorte que le système cognitif n'est plus considéré comme un système centralisé « dans la tête » d'un individu, mais comme un

système plus large, un système distribué (Zhang & Norman, 1994). Les conceptions dépendent donc, en partie, d'un système distribué qui les génère. Plusieurs auteurs (Leontiev, 1976 ; Balacheff, 1995 ; Leff, 2005) soutiennent que la conception que se fait un individu découle d'un processus de construction de connaissance soutenue par les interactions entre le sujet et son environnement physique et social. Ces interactions supposent l'existence d'une multitude de ressources génératrices des savoirs. C'est l'idée même de l'intelligence distribuée soutenue par la théorie de la « cognition distribuée ».

Cette idée est également soutenue par Balacheff (1995), dans ses travaux relatifs à l'étude des phénomènes de changement de « conception » des élèves en sciences à partir de la dynamique du système sujet/milieu. Balacheff postule donc 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 appelle donc par *conception* C, un quadruplet (P, R, L, Σ) 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 ;
- Σ est une structure de contrôle, elle assure la non-contradiction de C.

Cette acception de Balacheff est issue du paradigme de l'erreur et de la problématique des obstacles. Le paradigme de l'erreur (Brousseau, 1976), soutient que la trajectoire de l'élève devrait passer par la construction (provisoire) de connaissances erronées parce que la prise de conscience de ce caractère erroné serait constitutive du sens de la connaissance dont la construction est visée. Les points de passage obligés, l'auteur les nomme, à la suite de Bachelard (1938), « obstacles épistémologiques » : une connaissance, comme un obstacle, est toujours le fruit d'une interaction de l'élève avec son milieu et plus précisément avec une situation qui rend cette connaissance intéressante. Balacheff renonce ainsi au paradigme des « misconception » développé par Confrey (1986) et d'autres chercheurs. La perspective de Balacheff offre alors à notre étude des éléments théoriques nécessaires pour opérer ses choix méthodologiques et pour analyser ses résultats.

5 Choix méthodologiques

5.1 Choix des échantillons

Les critères de choix des échantillons sont dictés par la possibilité de leur comparabilité et par l'éventualité d'existence de conceptions en sciences contextuellement valides au regard de la différence des habitudes de vie pour les jeunes habitants des pays de continents assez éloignés. Deux échantillons indépendants du fait de leur langue d'enseignement et des habitudes de vie sont choisis dans deux pays des continents très éloignés. Les deux pays à échantillonner doivent participer à une même enquête internationale et que leur classement aux différentes sessions présente des écarts de performance récurrente. Nous avons donc choisi, sur la base de notre accessibilité aux items utilisés, les pays membres des enquêtes TEIMS. Notre choix a été particulièrement porté sur un pays maghrébin (le Maroc) et un pays nord-américain (Le Canada, plus particulièrement la province du Québec) parce que : -ces deux pays participent aux TEIMS ; -leurs contextes géographiques imposent des habitudes de vie différentes ; -leur classement aux différentes sessions de TEIMS en sciences présente une récurrence des écarts de performance (les valeurs chiffrées sont indiquées au tableau 1).

Par conséquent, les sujets à l'étude sont des jeunes marocains et québécois âgés de 15 ans environ. Cet âge correspond à la fin de la scolarité obligatoire pour les jeunes québécois et marocains. En définitive, un échantillon de 90 élèves de chaque pays a été retenu.

Les caractéristiques des 90 élèves recrutés sont :- être natif et originaire du pays ; - y avoir passé toute sa scolarité obligatoire ; habité un quartier où les traditions du pays orientent la vie quotidienne. Pour couvrir ces caractéristiques, nous avons procédé à la diversification des lieux de recrutement en fonction des spécificités culturelles des habitants. Au Maroc, les élèves ont été recrutés dans divers collèges de la ville traditionnelle de Salé. Au Québec, les élèves ont été recrutés dans diverses écoles secondaires de municipalités de la Rive-Sud de Montréal. La taille 90 élèves par pays permet de se faire une idée des tendances réponses dominantes pour des items qui présentent une importante différence de performance entre les deux pays.

5.2 Instrumentation de l'étude quantitative

Le but de notre étude quantitative est de doter notre recherche d'une base des données empiriques pouvant servir à identifier un nombre fini et facilement manipulable d'items (dix items environ) présentant les plus importantes différences de performance entre les deux échantillons. Dans l'ensemble de cette étude, l'item est considéré comme une question ayant des éventualités de choix de réponse bonne ou fausse. Les items utilisés dans cette étude ont été tirés des bases de données rendues publiques par les évaluations internationales TEIMS 2007 et PISA 2006. Ce choix est dû au fait que ces items sont déjà stabilisés et standardisés donc applicables pour évaluer un jeune de 15 ans environ de n'importe quel pays au monde. Un questionnaire de 20 items a été constitué pour évaluer la culture scientifique et non les connaissances scolaires. Cette nuance permet de contourner les difficultés dues au caractère national des programmes de formation. Les items ont été choisis également en fonction des concepts scientifiques dont la littérature en didactique des sciences dispose déjà des informations sur les conceptions véhiculées par des élèves.

Ces items impliquent deux domaines cognitifs : la connaissance et l'application. Le domaine de raisonnement n'a pas été retenu afin de faciliter la cotation et la comparabilité des réponses. Par conséquent, la double question à choix multiple (Coppens & Munier, 2005) a été retenue. Cette double question à choix multiple est constituée d'une partie évaluant la culture scientifique et une autre évaluant le degré de certitude exprimé (Caleon, & Subramaniam, 2010).

5.3 Exemple d'item

Pour faire la pâte à pain, 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.

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. Pourquoi la pâte monte-t-elle? (Souligne 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.



Figure 1: le pain

D) La pâte monte parce que la fermentation transforme l'eau en vapeur.

Es-tu sûr de ta réponse ? (Coche une seule case)

- Pas du tout sûr
- Pas très sûr
- Légèrement sûr
- Très sûr
- Extrêmement sûr

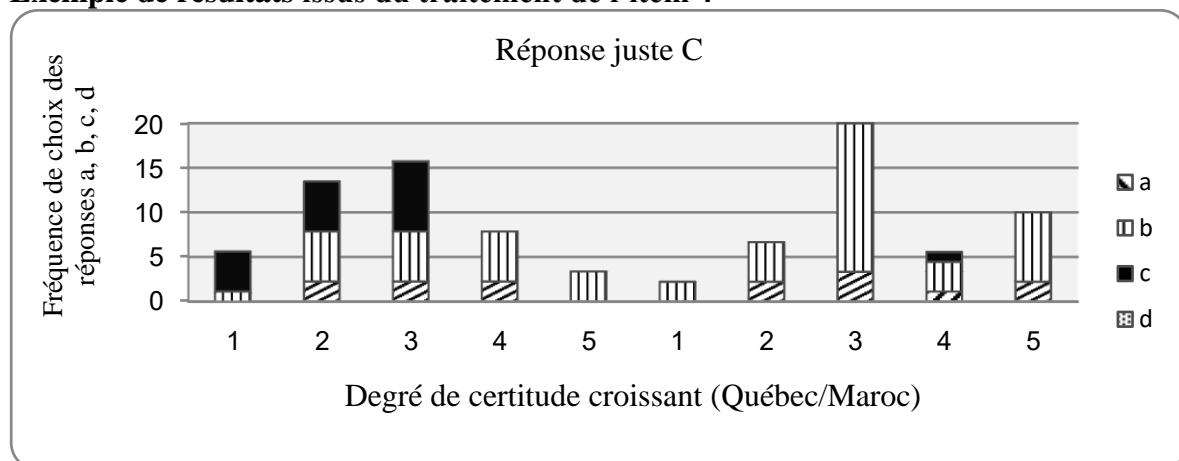
5.3 Procédé de traitement des données quantitatives

L'exploitation des données a été faite par un tri des réponses bonnes ou mauvaises en procédant par leur classification selon le degré de certitude exprimée. À l'image des travaux de Malhotra, Décaudin, et Bouguerra, (2007), notre échelle d'estimation de degré de certitude se présente comme suit :

Échelle par ordre croissant de degré de certitude

1	2	3	4	5
Pas du tout sûr	Pas très sûr	Légèrement sûr	Très sûr	Extrêmement sûr

Exemple de résultats issus du traitement de l'item 4



Ce graphique juxtapose les données de deux pays : en abscisse les degrés de certitude en ordre croissant de 1 à 5, en ordonnée les fréquences de choix des réponses a, b, c et d. Le graphique montre que pour tous degrés de certitude confondus, les élèves québécois ont choisi majoritairement la réponse c tandis que les élèves marocains ont choisi majoritairement la réponse b. Les réponses dominantes de part et d'autre indiquent une certitude légèrement sûre. Il y a donc nécessité d'aller chercher à comprendre les raisons qui motivent ces choix. C'est donc le but de l'étude qualitative.

5.4 Instrumentation de l'étude qualitative

La méthode d'entretiens individuels ou « interviews » utilisée dans notre étude qualitative a été adaptée des travaux de Tiberghien, Delacote & Guesde (1978) qui ont combiné à la fois l'entretien non directif et l'entretien directif pour étudier : - le niveau d'interprétation des élèves et leurs manières d'anticiper les phénomènes physiques ; - les relations entre le langage que les élèves utilisent et les notions que ce langage couvre ; - et les expériences personnelles, y compris dans le domaine affectif, auxquelles les élèves font référence. Leur approche est fondée sur : -une question tenant en une phrase avec un ou deux mots clefs introduisant la notion étudiée. - une situation expérimentale nécessitant soit une interprétation, soit une prévision. Pour la présente étude les situations expérimentales physiquement vécues ne font pas parti des enquêtes internationales donc cet aspect est exclu.

Nous réalisons nos entretiens en fonction et au regard des balises P, R, L, Σ de notre cadre théorique. Le tableau ci-après donne un aperçu des questions posées lors de l'entretien.

Tableau 2 : guide d'entretien individuel

Caractéristique d'une conception selon Balacheff	Question du chercheur
P ensemble de problèmes	Dis avec tes propres mots comment est-ce que tu comprends le problème posé ?
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	Est-ce que tes réponses proviennent de ce que tu as appris à l'école, de ta propre façon de voir les choses ou d'ailleurs ?
Σ 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?

Ces rubriques amènent donc l'élève à parler et à livrer sa propre conception de la tâche d'évaluation. Les entretiens sont conduits par le chercheur à l'aide d'un enregistreur audio et d'un journal de bord (Consignation des passages pour une relance ou une nouvelle hypothèse). La durée de l'entretien face à face est d'une (1) heure par élève, environ.

5.5 Modalités de traitement et d'analyse des données qualitatives

Notre étude opte pour une technique d'analyse systématique de contenu adaptée des travaux de Van der Maren (2003). La technique adaptée consiste à :

- 1) La préparation de l'analyse : Une grille d'analyse des données est élaborée sur la base des concepts scientifiques qui découleront des dix items présentant une différence significative de performance ;
- 2) L'analyse des traces : La détermination des unités d'analyse (mot ou concept) est faite au regard des concepts retenus. Il s'en suit une extraction des passages significatifs dans les propos retranscrits puis un codage suivi d'un classement selon les ressemblances selon les unités de sens ;
- 3) L'analyse de la qualité des données : Les données codées sont saisies puis vérifiées par la suite ;
- 4) La synthèse des données : Une condensation et une réduction des données par unités de sens codées est faite en utilisant le logiciel Winmax ;
- 5) La vérification et l'interprétation : Le chercheur procède à la vérification des effets des biais et fait contrôler les interprétations par les experts (triangulation des évaluateurs).

5.6 Principe de codage des verbatim

Premier niveau de codage : les passages significatifs

Les passages significatifs sont les phrases ou les segments de phrases qui présentent l'idée maîtresse ou l'essentiel de la réponse. Les passages significatifs sont regroupés par catégories et codifié en couleur. Le choix d'un code couleur est fait pour faciliter la classification des réponses par catégorie. Les catégories de passage significatif ci-après ont été retenues sur la base des indicateurs qui ont émergé à la suite d'un vas et viens entre les objectifs, les orientations du cadre théorique et les différentes réponses obtenues. En rouge : Compréhension du problème ; en bleu : Réponse exprimée ; en vert : Source de la réponse.

Deuxième niveau de codage : les unités de sens

Les unités de sens retenus sont les mots ou leurs synonymes ou les concepts qui marquent une signification spécifique. Les unités de sens sont soulignées dans les phrases des passages significatifs. Les unités de sens permettent le comptage afin de déterminer les fréquences de la redondance d'une idée au sein d'une catégorie de passage significatif. Par conséquent, dans la catégorie « compréhension du problème » les unités de sens sont les concepts, les mots ou leurs synonymes qui marquent un savoir ou une pratique de référence. Dans la catégorie « réponse exprimée » les unités de sens sont les concepts, les mots ou leurs synonymes qui marquent une connaissance contextuelle qu'elle soit scientifique ou naïve ou un doute à propos des choix de réponses proposées. Dans la catégorie « source de la réponse » les unités de sens sont les concepts, les mots ou leurs synonymes qui marquent le contexte d'émergence de la connaissance. Ce contexte peut être culturel ou scolaire.

5.7 Exemple de résultat obtenu à propos de l'item 4

Exemple de réponses obtenues durant la pré-enquête au Québec

AQ1 Olivier (nom fictif) tu vas lire cette question. Tu me dis comment est-ce que tu comprends la question. Lorsque tu avais répondu à cette question tu étais extrêmement sûr. *Oui !*

Ta réponse à la question était la pâte monte à cause des champignons unicellulaires qui se reproduisent. Est-ce que tu peux expliquer comment est-ce que tu as compris la question ?

A mon souvenir j'avais déjà vu un documentaire sur ça, comme justement dans la manière où la pâte était faite. Dans le documentaire on disait que les champignons unicellulaires se reproduisaient ce qui permettait la chaleur de se produire permettant ainsi la pâte de monter. Ta réponse vient du fait que tu as lu un documentaire et non tu as étudié cela en classe. Ouais !

Si on te proposait par exemple une telle réponse : la pâte monte parce qu'il y a un gaz le dioxyde de carbone qui se reproduit qu'on pense-tu ? *En moi ce pas ça mais comme ça se peut ça n'est pas vraiment impossible. Est-ce que tu as des exemples dans ta vie de tout le jour en dehors du documentaire qui peuvent t'aider à compléter ta réponse. Pas vraiment, le pain par exemple. Tu ne vas pas quelquefois à la boulangerie ou vous ne fabriquez pas du pain à la maison. Non d'habitude je trouve toujours le pain sur la table.*

CQ1 Ta réponse à cette question tu étais légèrement sûre. Ta réponse était que la pâte monte à cause à cause des champignons unicellulaire qui se reproduisent. Comment as-tu compris la question? *Je pense à une autre réponse. Je pensai que les champignons dans le fond faisaient en sorte que ça s'accumule dans le pain. Ça fait grossir le pain comme si ça faisait d'autres pains. Là je pense que c'est le dioxyde de carbone qui fait grossir. Est-ce que ta réponse c'est par rapport à ce que tu as étudié en classe. Non, c'est juste dans le fond on a étudié que le dioxyde de carbone est un gaz. Dans le fond un gaz fait toujours grossir les choses. Finalement as-tu quelques exemples qui appuient ta réponse comme telle. Mettant comme une ballonne d'air, si tu l'as gonfle avec du gaz elle va grossir. Alors pour toi c'est que la pâte monte parce qu'il y a un gaz à l'intérieur. Oui. Est-ce que chez vous vous avez l'habitude de fabriquer du pain à la maison ? Non. Vous achetez du pain où ? À l'épicerie. Tu n'étais jamais rendu à la boulangerie. Ben de fois pour aller acheter des gâteaux mais je ne regardais pas comment il faisait.*

A la suite de l'ajustement des questions du guide d'entretiens, une formulation unique a été adoptée afin d'obtenir le maximum d'information sans que chercheur pose plusieurs questions à l'élève. Car l'élève doit répondre à huit items en lui posant les mêmes questions. Cet afflux

des questions fatigue l'élève qui du reste termine en répondant par des gestes. De même, nous avons constaté qu'il suffisait de demander à l'élève ce qu'il pense du problème pour qu'il se mette à relater sa compréhension, ses difficultés, sa réponse préférée. Il ne manquait souvent que l'origine de la réponse et les liens avec son contexte de vie le cas échéant. Par conséquent, les questions ont été reformulées telles que présentées dans les exemples de réponses suivantes.

IQ1 Exemple de la réponse d'un élève québécois ayant fait le bon choix

Dis avec tes propres mots comment est-ce que tu comprends le problème posé ? *Pendant la fermentation, le pain subit une forte transformation chimique. Il y a la levure qui aide à transformer l'amidon, il y a aussi du sucre, la farine. Ça devient le dioxyde de carbone puis ça devient de l'alcool. Quelle réponse proposes-tu à la question posée ? Je ne suis pas trop sûr de ma réponse. J'ai pris comme je pense. J'ai pris le c). Le c) dit quoi ? La pâte monte parce qu'un gaz le dioxyde de carbone est produit. Qu'est-ce qui te rassure que c'est ça la réponse ? Je n'en sais rien. Juste je lis ce qui est décrit et j'essaie de penser ce qui arrive. Est-ce que tu as étudié ça en classe ? La fermentation peut être mais je ne suis pas certain.*

Exemple de la réponse d'un élève québécois ayant fait le même choix que la majorité des élèves marocains.

Dis avec tes propres mots comment tu comprends le problème posé ? *J'ai compris que la question à un rapport avec la chimie. Les réactions chimiques et les mélanges. Ça un rapport avec les réactions chimiques qui se passent après. La fermentation fait monter la pâte. Pourquoi la pâte monte-t-elle ? Quelle serait la réponse selon toi qui conviendrait à la question ? Alors puisque la levure est le champignon unicellulaire aide à transformer l'amidon et le sucre de la farine en dioxyde de carbone genre en gaz et en alcool. Ça veut dire que si on prend la première parce que l'alcool produit se transforme en gaz. L'alcool ne se transforme pas en gaz. Parce que c'est déjà l'amidon qui est transformé en dioxyde de carbone puis en alcool. Donc la première ce n'est pas une bonne réponse. La pâte monte à cause des champignons unicellulaires qui se reproduisent. Peut-être. La pâte monte parce que la fermentation transforme l'eau en vapeur. L'eau ne se transforme pas en vapeur dans ce cas-ci. Est-ce que ta réponse a un lien avec ce que tu as appris à l'école ? Moi je prends ça par rapport à ce que je comprends par rapport à ma logique à moi. C'est ce que je comprends de la question. Alors la pâte monte à cause des champignons unicellulaires qui se reproduisent. Puis on ne peut pas dire que la pâte monte à cause d'un gaz puisqu'il y a d'autres choses qui se reproduisent à cause de champignon unicellulaire. La pâte est déjà reproduite à cause des champignons unicellulaires mais il y a aussi d'autres comme l'alcool qui se reproduit. Fait qu'on peut unifier tout ça. Dans la réponse la pâte monte à cause des champignons unicellulaires qui se produisent.*

Exemple de réponse d'un élève Marocain ayant fait le mauvais choix dominant.

AM1 Dis avec tes propres mots comment tu comprends le problème posé ? *Moi, je dirai le b) la pâte monte à cause des champignons unicellulaires qui se reproduisent. D'abord comment est-ce que tu as compris le problème ? C'est que pour faire du pain. Il faut de la farine et de la levure. Et qu'après la pâte elle est placée quelque part dans un conteneur pendant plusieurs heures pour aider qu'il gonfle. Pendant qu'il gonfle, il y a un changement. La levure aide a transformé les autres ingrédients pour que le pain il monte. Est-ce que ta réponse c'est par rapport à ce que tu as appris à l'école ? Non pas à l'école. Comment est-ce que tu as trouvé cette réponse ? Ben plus à mon expérience à faire du pain chez moi. Fait qu'avec ma mère je fais du pain puis elle me chicane, il faut que ça monte. Puis je regardai comment est-ce ça monte puis elle me dit c'est à cause de la levure.*

6 Discussion et conclusion

Plusieurs exemples de la sorte ont été recueillis. Nous disposons d'une banque des données québécoises et marocaines de 112 verbatim pour les 7 items étudiés. Les résultats complets ne sont pas encore tous analysés. Cependant nous offrons ici à notre lecteur un aperçu sur l'existence des données empiriques qui permettent d'interroger le caractère standard des items des évaluations internationales en sciences. Les élèves québécois en majorité dans l'exemple cité se sont inspirés des données du texte pour y chercher la bonne réponse. Dans leur vie quotidienne le pain est fabriqué loin des familles notamment dans les boulangeries auxquelles ils n'ont accès qu'au comptoir de vente ou à l'étalage de l'épicerie. Le contexte culturel ne leur permet pas de répondre avec beaucoup de certitude à un problème concernant la fabrication du pain. Nombreux ont donc avoué ne pas être sûr de leur réponse bien que juste.

Pour les élèves marocains, dans leur vie familiale de tout le jour la fabrication du pain fait partie de leurs habitudes traditionnelles. Ils utilisent la levure pour faire monter la pâte. Cette conviction totalement ancrée dans leur contexte culturel leur amène à exploiter certes le texte comme les élèves québécois. Cependant, tout en étant convaincu que dans le texte on parle de la levure qui est un champignon unicellulaire et que l'une de réponse proposée parle du champignon unicellulaire qui se reproduit pour faire monter la pâte, les élèves marocains ont choisis majoritairement cette réponse. Au regard donc de ce qu'ils savent déjà en pratique, cette réponse leur paraît juste. La question est de savoir si la performance des élèves marocains à cette question est démeritée?

L'item ci-dessus présenté a déjà été utilisé en 2006 par PISA pour évaluer puis classer les pays asiatiques, américains, européens, arabes de l'Afrique du nord et du moyen orient. Cet item a donc été stabilisé et validé par des techniques éduométriques éprouvées (Bertrand & Blais, 2004). Par conséquent, l'explication des différences de performance aux évaluations internationales entre les pays sur la base uniquement des données quantitatives est confrontée aux conceptions d'origine culturelle qui sont derrière les réponses contextuellement valides. Dans cette étude, les conceptions d'origine culturelle identifiées et la totalité des données seront rendues publiques prochainement. Toutefois, une question émerge de cette étude : Quels peuvent être les invariants dans l'habillage d'un item dit standardisé en sciences au regard des conceptions d'origine culturelle qui orientent les réponses des élèves?

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* (p.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*. Nicosia: Cyprus University Press, p. 108-123.
- 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*. London: Routledge, p. 277-290.
- Bachelard, G. (1938). *La formation de l'esprit scientifique*. Paris : Vrin.
- Balacheff, N. (1995). Conception, propriété du système sujet/milieu. In R. Noirfalise., M.-J. Perrin-Glorian (eds.) *Actes de la VII^e École d'été de didactique des mathématiques* (p.215-229). Clermont-Ferrand: IREM de Clermont-Ferrand.
- Bertrand, R., & Blais, J.G. (2004). *Modèles de mesure – L'apport de la théorie des réponses aux items*. Sainte-Foy, Québec : Presses de l'Université du Québec.

- Blum, A., Goldstein H. & Guerin-Pace, F. (2001). International adult literacy survey (IALS): an analysis of international comparisons of adult literacy. *Assessment in Education*, vol. 8, n° 2, p. 225-246.
- Boutin, G. (2000). *L'entretien de recherche qualitatif*. Sainte-Foy : Presses de l'Université du Québec.
- Brousseau, G. (1976). Les obstacles épistémologiques et les problèmes en mathématiques. In (1983) *Recherches en didactique des mathématiques*, 4(2), 164-198.
- Bulles, N (2010, mars-avril). L'imaginaire réformateur. PISA et les politiques de l'École. *Le Débat*, n°159, p.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.
- Carr, M. (1996). *Interviews about instances and interviews about events*. New York: Teachers College Press.
- 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.
- Chi, M. T. H. (2008). Three types of conceptual change: belief revision, mental model transformation and categorical shift. In S. Vosniadou (Dir.), *International Handbook of research of conceptual change* (p. 61-82). New-York: Routledge.
- Clarke, D. (2006). The LPS research design. In D. Clarke, C. Keitel & Y. Shimizu (Eds.) *Mathematics in twelve countries: the insider's perspective*, pp. 15-36. Rotterdam: Sense Publishers.
- Confrey, J. (1986). Misconceptions accross sujet matters : charting the course from a constructivist perspective. *Annual meeting of the American Educational Research Association*. Document photocopié.
- Coppens, N., & Munier, V. (2005). Évaluation d'un outil méthodologique, le double QCM, pour le recueil de conceptions et l'analyse de raisonnements en physique. *Didaskalia*, 27, 41-77.
- Creswell, J.W. (1994). *Research design Qualitative and quantitative approaches*. Thousand Oaks/London/New Delhi: SAGE Publications.
- Duit, R. (1999). Conceptual Change Approaches in Science Education. In W. Schnotz., S. Vosniadou., & M. Carretero (Eds.), *New Perspectives on Conceptual Change* (p.263-282). Amsterdam: Pergamon Press.
- Duit, R., & Treagust D.F. (2003). Conceptual change: a powerful framework for improving science teaching and learning. *International Journal of Science Education*, 25 (6), 671-688.
- Finegold, M., & Gorsky, R (1991). Students' concepts of force as applied to related physical systems: a search for consistency. *International Journal of Science Education*, 13(1), 97-113.
- Franklin, B. J. (1992). The development and application of a two-tier diagnostic instrument to detect misconceptions in the area of force, heat light and electricity. *Dissertation Abstracts International*, 53(12), 41-86,
- Gilles, J.-L. (2002). Qualité spectrale des tests standardisés universitaires – Mise au point d'indices éducatifs d'analyse de la qualité spectrale des évaluations des acquis des étudiants universitaires et application aux épreuves MOHICAN check up 99. Thèse de doctorat inédite en Sciences de l'éducation. Liège : Université de Liège, Faculté de psychologie et des sciences de l'éducation de l'université de Liège.
- Givry, D. (2003). Le concept de masse en physique : quelques pistes à propos des conceptions et des obstacles. *Didaskalia*, 22, 41-6
- Goldstein, H. (2004b). International comparative assessment: how far have we really come? *Assessment in Education*, vol. 11, n° 2, p. 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, vol. 1, p. 99-112.
- Gustafsson, J.-E. & Rosen, M. (2004). « The IEA 10-Year Trend Study of Reading Literacy: A multivariate reanalysis ». In C. Papanastasiou (éd.), *Proceedings of the IRC-2004*. Nicosia: Cyprus University Press, vol. 1, p. 99-112.
- Hallden, O. (1999). Conceptual change and contextualisation. In W. Schnotz & S. Vosniadou & M. Carretero (Eds.), *New Perspectives on Conceptual Change* (p. 53-65). Amsterdam: Pergamon Press.

- Halloun, I. A., & Hestenes, D. (1985). The initial knowledge of college physics students. *American Journal of Physics*, 53(11), 1043-1055.
- Harlow, A. & Jones, A. (2004). « Why Students Answer TIMSS Science Test Items the Way They Do ». *Research in Science Education*, vol. 34, n° 2, p. 221-238.
- Hegarty-Hazel, E., & Prosser, M. (1991). Relationship between students' conceptual knowledge and study strategies. Part 1: Student learning in physics. *International Journal of Science Education*, 13, 303-312.
- Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.
- Hutchison, G., & Schagen, I. (2007). Comparisons between PISA and TIMSS – Are We the Man with Two Watches? In Loveless, T. (Ed.), *Lessons Learned – What International Assessments Tell Us about Math Achievement*. Washington, DC: The Brookings Institution.
- Joshua, S., & Dupin, J.-J. (1993). *Introduction à la didactique des sciences et des mathématiques*. Paris: PUF.
- 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.
- Kyriadkides, L., & Charalambous, C. (2005). Using educational effectiveness research to design international comparative studies, *Research Papers in Education*, 20(4), 391-412.
- Lafontaine, D. (2010). Les standards internationaux en éducation : la place des situations et des contextes. In D. Masciotra, F. Medzo et Ph. Jonnaert, *Vers une approche située en éducation : Réflexions, pratiques, recherches et standards* (p.159-174). *Cahier scientifique 111*, Acfas.
- Lave, J. (1988). *Cognition in Practice: Mind, mathematics, and culture in everyday life*. Cambridge, UK: Cambridge University Press.
- Leduc, D., Riopel, M., Raïche, G. et Blais, J.-G. (2011). L'influence des définitions des habiletés disciplinaires sur la création et le choix d'items dans le PISA et le TEIMS. *Mesure et évaluation en éducation*, 34 (1), 97-130.
- Leontiev, A. (1974). The problem of activity in psychology. *Soviet psychology*, 13(2), 4-33.
- Lee, G., Kwon, J., Park, S.-S., & Kim, J. W. (2003). Development of an instrument for measuring cognitive conflict in secondary level science classe. *Journal of research in science teaching*, 40(6), 585-603.
- Lundeberg, M. A., Fox, P. W., Brown, A. C., & Elbedour, S. (2000). Cultural influences on confidence: country and gender. *Journal of Educational Psychology*, 92(1), 152-159.
- Malhotra, N. K., Décaudin, J.-M., & Bouguerra, A., (2007). *Etudes marketing avec SPSS*. 5^{ème} éd, Pearson Education.
- Messick, S. (1994). *Alternative modes of assessment, uniform standards of validity*. Princeton, NJ: Educational Testing Service.
- Millar, R., & Hames, V. (2001). Using diagnostic assessment to improve students' learning: some preliminary findings from work to develop and test diagnostic tools. In D. Psillos & al., (Eds). *Proceedings of the Third International Conference on Science Education Research in the Knowledge Based Society* (p 141-143). Thessaloniki, Greece: Art of text Publications.
- Meunier, O. (2005). Standards, compétences de base et socle commun. *Dossier de synthèse, veille scientifique et technologique*, Lyon : INRP récupéré de <http://www.inrp.fr/vst/Dossiers/Standards/sommaire.htm> le 25-12 2010.
- Mons, N. (2008). Evaluation des politiques éducatives et comparaisons internationales : introduction. *Revue française de pédagogie*, n° 164, p.5-13.
- Nadelson, L. S., & Southerland, S. A. (2010). Development and preliminary evaluation of the Measure of Understanding of Macroevolution: Introducing the MUM. *The Journal of Experimental Education*, 78, 151-190.

- Nandakumar, R. (1994). Assessing dimensionality of a set of item responses- Comparison of different approaches. *Journal of Educational Measurement*, 31, 17-35.
- Neidorf, T.S., Binkley, M., Gattis, K., & Nohara, D. (2006). Comparing Mathematics Content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 Assessments (NCES 2006-029). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Palmer, D. (1997). The effect of context on students' reasoning about forces. *International Journal of Science Education*, 19(6), 681-696.
- 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.
- Potvin, P., Riopel, M., Charland, P., Mercier, J. (2011). Portrait des différences entre les genres dans le contexte de l'apprentissage de l'électricité en fonction de la certitude exprimée lors de la production de réponses. *Canadian Journal of science, mathematics and technology education*, 11(4), 328-347.
- Rolf V. Olsen, Svein Lie (2006). Les évaluations internationales et la recherche en éducation : principaux objectifs et perspectives. *Revue française de pédagogie*, n° 157, octobre-novembre-décembre, 11-26.
- Säljö, R. (1999). Concepts, cognition and discourse: from mental structures to discursive tools. In W. Schnotz & S. Vosniadou & M. Carretero (Eds.), *New Perspectives on Conceptual Change* (pp. 53-65). Amsterdam: Pergamon Press.
- Shealy, R. T., & Stout W. F. (1993). « A model-based standardization approach that separates true bias/DIF from group ability differences and detect test bias/DTF as well as item bias/DIF ». *Psychometrika*, 58, 159-194.
- Thijs, G. D., & Van Den Berg, ED. (1995). Cultural Factors in the Origin and Remediation of Alternative Conceptions in Physics. *Science & Education*, 4, 317-347.
- Tiberghien, A., Delacote, G., & Guesde, E. (1978, oct, nov, déc). Méthodes et résultats concernant l'analyse des conceptions des élèves dans différents domaines de la physique : Deux exemples : les notions de chaleur et lumière. *Revue française de pédagogie*. 45, 25-32.
- Treagust, D. F., & Duit, R. (2008). Compatibility between cultural studies and conceptual change in science education: there is more to acknowledge than to fight straw men! *Cultural Studies of Science Education*, 3, 387-395.
- Tsai, C.-C., & Chou, C. (2002). Diagnosing students' alternative conceptions in science. *Journal of Computer Assisted Learning: Blackwell Science*, 18(2), 157-165.
- Tsui, C.Y., & Treagust, D. (2009). Evaluating Secondary Students' Scientific Reasoning in Genetics Using a Two-Tier Diagnostic Instrument. *International Journal of Science Education*, 1-26.
- 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 Education, Noordwijkerhout [Pays-Bas].
- Tüysüz, C. (2009, juin). Development of two-tier diagnostic instrument and assess students' understanding in chemistry. *Scientific Research and Essay*, 4 (6), 626-631.
- Van der Maren, J.-M. (2003). *La recherche appliquée en pédagogie : Des modèles pour l'enseignement 2^{ème} éd.* Bruxelles : De Boeck Université.
- Viennot, L. (1977). Le raisonnement spontané en dynamique élémentaire. Thèse de doctorat d'État inédite. Paris : université Paris 7.
- Vrignaud, P. (2002). Les biais de mesure : savoir les identifier pour y remédier. *Bulletin de psychologie*, 55(6), 625-634.
- 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.
- Wu, M. L. (2008, March). A Comparison of PISA and TIMSS 2003 achievement results in Mathematics. Paper presented at the AERA Annual Meeting, New York.
- Zhang, J., & Norman, D. A. (1994). Representations in Distributed Cognitive Tasks. *Cognitive Science*, 18(1), 87-122.