



uOttawa

L'Université canadienne
Canada's university

**FACULTÉ DES ÉTUDES SUPÉRIEURES
ET POSTDOCTORALES**



uOttawa

L'Université canadienne
Canada's university

**FACULTY OF GRADUATE AND
POSTDOCTORAL STUDIES**

Sarah Plouffe

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.A. (Education)

GRADE / DEGREE

Faculty of Education

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

**Analyse de la qualité psychométrique des items issus
des questionnaires contextuels de la Teims**

TITRE DE LA THÈSE / TITLE OF THESIS

Marielle Simon

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Renée Forgette-Giroux

Dany Laveault

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

**ANALYSE DE LA QUALITÉ PSYCHOMÉTRIQUE DES ITEMS
ISSUS DES QUESTIONNAIRES CONTEXTUELS DE LA TEIMS**

Sarah Plouffe

Thèse soumise à la
Faculté des études supérieures et postdoctorales
dans le cadre des exigences
du programme de maîtrise

Faculté d'éducation

Université d'Ottawa

© Sarah Plouffe, Ottawa, Canada, 2007



Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-49264-2
Our file Notre référence
ISBN: 978-0-494-49264-2

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Remerciements

« C'est le devoir de chaque homme de rendre au monde au moins autant qu'il en a reçu. »

Albert Einstein

Lorsque la vie nous sourit, elle nous offre parfois des opportunités ou des expériences qui surpassent toutes nos attentes. Ma première rencontre avec professeur Marielle Simon n'aurait pu me laisser présager une association aussi bénéfique. Dans la vie, les rencontres qui me sont les plus chers sont celles qui me permettent soit d'acquérir une perspective différente, d'accéder à de nouvelles connaissances ou encore de mieux connaître mes propres forces et faiblesses. Professeur Simon, par la sagesse de ces généreuses et bienveillantes interventions, a toujours su me guider vers la rencontre de ces trois sommets et je lui en est que reconnaissante. Elle a su me transmettre sa passion pour le domaine de l'éducation et bien sûr la piqûre de la mesure!

J'aimerais également témoigner ma reconnaissance envers les membres de mon comité, professeur Renée Forgette-Giroux et professeur Dany Laveault de l'université d'Ottawa. J'ai pu profiter de leur imposante réputation, de leur important bagage de connaissances ainsi que de leur vaste expérience dans le domaine. Leurs commentaires, toujours formulés de façon très constructive, m'ont été très précieux dans la rédaction de ce document.

Je n'aurais pu réaliser un tel projet sans l'aide et l'encouragement de ma famille et de mes amis. Je voudrais d'abord remercier mon père qui m'a transmis sa passion pour la lecture, sa soif de connaissances et ces belles valeurs de respect, d'émerveillement et de découverte devant les phénomènes de la nature. À ma mère, je dois tout. Je lui dois en vérité, toute une vie, puisqu'elle a voué la sienne à mon bonheur et à mon épanouissement. Je remercie également mon frère de ses encouragements, de ses silences et de sa façon de dire les choses sans vraiment les dire. Il a été mon premier sujet de recherche, peut-être sans le vouloir, et je lui dois mon choix de carrière dans le domaine de l'éducation.

Et finalement, je remercie mon noyau, mon île, ma richesse... Christopher, Alexandre et Thomas. À toi Christopher, qui a toujours eu les bonnes paroles pour m'encourager, la patience d'un moine pour m'écouter, les talents d'un informaticien-expert pour retrouver une thèse supprimée, la charité des soupers préparés, des bébés bercés, des tâches ménagères exécutées pour me laisser travailler. Mais surtout, je te remercie d'avoir cru en moi lorsque j'avais peine à le faire pour moi-même et de m'avoir fait comprendre, sans même l'articuler, que la poursuite de mes rêves était une priorité! Merci à Alexandre et Thomas à qui je dois encore quelques histoires avant le coucher et bien des journées uniquement consacrées à s'amuser. Vous me donnez à tous les jours le courage de continuer et surtout le goût de rêver encore plus grand!

Résumé

De nos jours, les dirigeants consultent les résultats de programmes nationaux et internationaux d'évaluation des apprentissages afin de se renseigner au sujet de la qualité de l'enseignement dispensé dans les salles de classe. Cependant, des études ont remis en question la qualité des données issues de ces questionnaires contextuels au sujet des pratiques et croyances pédagogiques (Clare, 2000; Ball & Rowan, 2004, Hill, 2002; Mayer 1999, Mullens & Kasprzvk, 1999; Mullis, 2002; NCES, 1999; NCES, 2002; Porter & Smithson, 2001; Rogers et al., 2004; Rowan, Correnti & Miller, 2002; Stecher et al. 2002; Stigler & Hiebert, 1996). La présente recherche procède à l'évaluation de la qualité psychométrique des items issus des questionnaires contextuels de la Troisième enquête sur les mathématiques et les sciences (TEIMS) portant sur les pratiques pédagogiques recommandées par les réformes scolaires. L'étude cible la population des enseignants francophones de mathématiques oeuvrant en Ontario. Une méthodologie mixte est employée pour mesurer divers concepts de validité et de fidélité des items de la TEIMS.

Les lacunes relevées au niveau de la validité proviennent principalement de la conception des items (formulation de la question, vocabulaire utilisé, période de référence de l'item, pertinence du sujet et des choix de réponse). La concordance entre élève et enseignant démontre que les items simples produisent des résultats davantage comparables que les items complexes faisant appel à des concepts vagues ou trop nombreux. La validité théorique n'a pu être établie en fonction des regroupements conceptuels inspirés par le curriculum de l'Ontario (1997). Les propriétés métriques des

items s'avèrent également problématiques compte tenu du type de données générées et des lacunes distributives des items. Toutefois, les analyses n'ont pas permis de déceler la présence d'un biais de réponse systématique en fonction du phénomène de désirabilité sociale. Les analyses de cohérence interne démontrent une faible cohésion entre les items des questionnaires qui traitent des nouvelles pratiques et croyances pédagogiques.

Ces résultats indiquent en grande partie que les données provenant de questionnaires contextuels des programmes à grande échelle n'arrivent pas à dresser un portrait valide ni fidèle des nouvelles pratiques pédagogiques chez les enseignants. Les résultats de cette étude doivent cependant, être interprétés avec discernement à cause de certaines faiblesses méthodologiques, dont les données non-représentatives de la population ciblée et le choix limité des méthodologies.

Table des matières

INTRODUCTION	9
RECENSION DES ÉCRITS.....	12
<i>Introduction d'un nouveau curriculum en Ontario.....</i>	<i>12</i>
<i>Évaluation des systèmes d'éducation.....</i>	<i>13</i>
<i>Qualité psychométrique des questionnaires contextuels.....</i>	<i>16</i>
CADRE DE RÉFÉRENCE.....	31
<i>Cadre sous-jacent à la TEIMS.....</i>	<i>31</i>
<i>Cadre sous-jacent à la qualité psychométrique du questionnaire contextuel.....</i>	<i>34</i>
MÉTHODOLOGIE.....	44
<i>Participants.....</i>	<i>44</i>
<i>Instruments.....</i>	<i>47</i>
<i>Questionnaires contextuels de la TEIMS.....</i>	<i>47</i>
<i>Questionnaire à l'intention du panel d'enseignants.....</i>	<i>50</i>
<i>Procédure.....</i>	<i>54</i>
<i>Traitement des données.....</i>	<i>55</i>
RÉSULTATS	63
<i>Caractéristiques démographiques des échantillons.....</i>	<i>63</i>
<i>Propriétés métriques des variables de l'étude.....</i>	<i>66</i>
<i>Validité d'apparence et de contenu.....</i>	<i>83</i>
<i>Concordance entre élève et enseignant.....</i>	<i>98</i>
<i>Validité théorique.....</i>	<i>101</i>
<i>Biais de réponse.....</i>	<i>102</i>
<i>Fidélité.....</i>	<i>103</i>
DISCUSSION.....	105
<i>Sommaire des produits de la recherche.....</i>	<i>105</i>
<i>Contextualisation des résultats de l'étude.....</i>	<i>114</i>
<i>Limites de l'étude.....</i>	<i>120</i>
CONCLUSION	122
RÉFÉRENCES	125
ANNEXES	134

Annexe A - Inventaire des items sélectionnés à partir des questionnaires	
Contextuels	135
Annexe B - Certificat d'approbation déontologique	138
Annexe C - Tableau des items sélectionnés à partir des questionnaires enseignants	140
Annexe D - Documents contenus dans la trousse acheminée au panel d'enseignants...	153
Annexe E - Résultats ANOVA	168
Annexe F - Résultats des analyses factorielles exploratoires	180
Annexe G - Tableaux des taux de réponse obtenus pour les items enseignants	184
Annexe H - Tableaux des coefficients d'asymétrie et de kurtose	188
Annexe I - Tests de distribution normale (Kolmogorov-Smirnov et Shapiro-Wilk)	192
Annexe J - Tableaux des mesures de dispersion	196

Liste des tableaux

Tableau 1 : Limites rapportées au sujet des items traitant des pratiques pédagogiques	17
Tableau 2 : Éléments constitutifs du cadre psychométrique	35
Tableau 3 : Nombre d'enseignants, de classes et d'élèves dans les trois échantillons	46
Tableau 4 : Nombre d'items retenus à partir des questionnaires enseignants et élèves en fonction des compétences pédagogiques ciblées	49
Tableau 5 : Questions adressées au panel d'experts	52
Tableau 6 : Concepts, critères et méthodes d'analyse	56
Tableau 7 : Répartition selon le sexe des enseignants dans les trois échantillons	64
Tableau 8 : Répartition selon l'âge des enseignants dans les trois échantillons	65
Tableau 9 : Années d'expérience des enseignants des trois échantillons	65
Tableau 10 : Répartition selon le niveau d'éducation des enseignants dans les trois échantillons	66
Tableau 11 : Pourcentage du type de variables générées par les questionnaires de la TEIMS	67
Tableau 12 : Moyenne, médiane et mode des items enseignants de l'échantillon de 1995	72
Tableau 13 : Moyenne, médiane et mode des items enseignants de l'échantillon de 1999	73
Tableau 14 : Moyenne, médiane et mode des items enseignants de l'échantillon de 2003	74
Tableau 15 : Items de la TEIMS présentant une asymétrie et une kurtose problématiques	82
Tableau 16 : Pourcentage d'items à chaque position occupée dans l'ordre des sous-questions	90
Tableau 17 : Résultats standardisés des déviations moyennes entre les élèves d'une même classe	99
Tableau 18 : Résultats standardisés des déviations moyennes entre les élèves et leur enseignant	101
Tableau 19 : Résultats de l'analyse de fidélité des indices composés pour les trois thèmes de l'étude	104

Liste des figures

Figure 1 : Interactions entre les trois niveaux de curriculum	33
Figure 2 : Taux de réponse minimum et maximum aux items du questionnaire enseignant – pour les trois cycles – données originales	69
Figure 3 : Taux de réponse minimum et maximum aux items du questionnaire enseignant – pour les trois cycles – données finales	70
Figure 4 : Illustration des bornes servant à identifier les items ayant des moyennes problématiques	71
Figure 5 : Étendues et écarts types des items de l'échantillon 1995	78
Figure 6 : Étendues et écarts types des items de l'échantillon 1999	79
Figure 7 : Étendues et écarts types des items de l'échantillon 2003	79
Figure 8 : Courbes des moyennes d'items démontrant des hausses significatives entre les cycles	93
Figure 9 : Courbes des moyenne d'items démontrant des baisses significatives entre les cycles	95
Figure 10 : Courbes des moyennes d'items ne démontrant pas de tendance linéaire	96

Introduction

Au cours des dernières décennies, plusieurs pays ont élaboré des politiques de réforme de leurs systèmes d'éducation en vue de former les élèves d'une nouvelle génération; l'objectif étant que ceux-ci puissent répondre aux besoins des marchés maintenant fondés sur le haut-savoir et les nouvelles technologies (O'Connor-Petruso, Schiering, Hayes & Serrano, 2004). L'élaboration de politiques de réforme et la mise en application de nouveaux curriculums visent avant tout, une amélioration de la qualité de l'enseignement pour les élèves en salle de classe. Or, il est clair que les enseignants sont les principaux agents de la mise en œuvre des politiques de réforme portant sur le contenu de la matière à enseigner et les pratiques pédagogiques à privilégier. En effet, de plus en plus d'études rapportent maintenant que les enseignants et la qualité de l'instruction qu'ils dispensent aux élèves semblent avoir un impact majeur sur le rendement de ces derniers (Loewenberg-Ball & Rowan, 2004; O'Connor-Petruso, Schiering, Hayes & Serrano, 2004). Cependant, la qualité des données disponibles actuellement au sujet de l'enseignement prodigué laisse planer un certain doute et les besoins se font apparents pour l'obtention d'une information sûre et entière afin de guider les futurs efforts de réforme. Le manque d'information provient du fait que, de par sa nature complexe, l'enseignement s'avère difficilement mesurable (Mayer, 1999) et que les techniques de l'heure n'arrivent pas à répondre à tous les besoins des preneurs de décisions en matière d'information.

De nos jours, les dirigeants consultent les résultats de programmes nationaux et internationaux d'évaluation des apprentissages afin de se renseigner au sujet de la qualité de l'enseignement dispensé dans les salles de classe. En plus d'évaluer le rendement et

certaines caractéristiques internes des élèves, ces programmes sondent les enseignants par l'entremise de questionnaires contextuels. La plupart des programmes d'évaluation recueillent de l'information sur les caractéristiques des enseignants (âge, sexe, expérience, scolarité, etc.), la matière enseignée, ainsi que les pratiques et les croyances pédagogiques adoptées par ces enseignants. Ces données font l'objet d'analyses statistiques primaires et secondaires dans le but d'informer les dirigeants et de les guider lors de l'élaboration de nouvelles politiques de réforme scolaire (Pelgrum, Voogt & Plomp, 1995; Puk, 1999). Cependant, au cours des dernières années, des études ont remis en question la qualité des données issues de ces questionnaires contextuels (Clare, 2000; Ball & Rowan, 2004, Hill, 2002; Mayer 1999, Mullens & Kasprzvk, 1999; Mullis, 2002; NCES, 1999; NCES, 2002; Porter & Smithson, 2001; Rogers et al., 2004; Rowan, Correnti & Miller, 2002; Stecher et al. 2002; Stigler & Hiebert, 1996). En effet, afin de mieux répondre à leurs besoins respectifs, le monde des réformes scolaires et celui des programmes d'évaluation se posent les questions suivantes : Est-ce que les données disponibles provenant des évaluations à grande échelle sont adéquates pour informer les décideurs au sujet de ce qui se passe véritablement dans les salles de classe ? Est-ce que le contenu des questionnaires est suffisamment représentatif des nouveaux programmes éducatifs des populations participantes, en particulier celles qui adoptent un renouveau pédagogique ? Est-ce que la qualité des instruments de collecte de données permet de faire confiance aux résultats obtenus lors des analyses ? Est-ce que le format des items permet de mener efficacement les analyses statistiques nécessaires pour informer les décideurs ? Comment les questionnaires contextuels de ces programmes d'évaluation peuvent-ils être améliorés pour mieux répondre aux besoins d'information ?

La présente recherche propose de cibler certaines de ces problématiques en évaluant la qualité psychométrique des items issus des questionnaires contextuels de la Troisième enquête sur les mathématiques et les sciences (TEIMS) portant sur les pratiques pédagogiques recommandées par les réformes scolaires. La prochaine section contient une recension des écrits sur 1) les enjeux introduits par le récent phénomène de la réforme scolaire, 2) les principales tendances en matière d'évaluation des systèmes d'éducation ainsi que 3) les différentes limites inhérentes aux données générées par les évaluations à grande échelle. Ensuite, le cadre de référence de l'étude est présenté accompagné de la question de recherche. La section suivante aborde les aspects méthodologiques de l'étude, notamment les populations étudiées, l'instrument, la procédure et les analyses. La prochaine section rapporte l'ensemble des résultats issus des analyses quantitatives et qualitatives réalisées autour des items des questionnaires contextuels de la TEIMS-R. Pour clore cet essai, une discussion sur la contribution de l'étude, les limites de celle-ci, ainsi que des pistes de recherches futures sont présentées.

Recension des écrits

La recension des écrits comprend deux sections distinctes, la première partie sert à établir le contexte de l'étude en présentant d'abord les grandes lignes du curriculum de mathématiques introduit en Ontario en 1997. Cette section contient également une brève description des programmes d'évaluation à grande échelle et leur importance dans un contexte de renouveau pédagogique. Elle se concentre en particulier sur les données directement liées aux croyances et aux pratiques pédagogiques des enseignants. La troisième section rapporte les études ayant critiqué la qualité psychométrique des questionnaires contextuels de programmes d'évaluation à grande échelle.

Introduction d'un nouveau curriculum en Ontario

Si les premiers élans de réforme scolaire avaient souvent comme objectif de rendre l'éducation plus accessible à l'ensemble de la population d'élèves d'âge primaire, il en est autrement de nos jours (OCDE, 1995). Les plus récentes politiques de réforme misent plutôt sur le principe de qualité de l'éducation puisque celle-ci exercerait un impact direct sur le rendement des élèves (Fuhrman, 2003). En effet, Blank et Pechman (1995) rapportent que les recommandations faites par les réformateurs sous forme de politiques éducatives portent maintenant, plus que jamais, sur les pratiques et les stratégies d'enseignement. Il s'agit d'une tendance souhaitable puisque, selon Schmidt et Cogan (1996), les pratiques pédagogiques que s'approprient les enseignants déterminent le niveau de la qualité de l'éducation.

Dans le domaine des didactiques mathématiques, les recommandations du *National Council of Mathematics* (NCTM 1989, 1991, 1995) ont, à ce jour, inspiré l'élaboration d'un

grand nombre de politiques éducatives à travers le monde (Burstein, 1995; Masini & Taylor, 2000; Mayer, 1999; Spillane & Zeuli, 1999). À titre d'exemple, en 1997, le Ministère de l'éducation et de la formation de l'Ontario (MEFO) effectue un remaniement important de son programme d'études à l'élémentaire en présentant un nouveau curriculum appliqué à l'ensemble de son territoire. Ce curriculum mise sur une nouvelle vision de l'apprentissage, de l'enseignement et de l'évaluation des mathématiques et témoigne directement de l'influence des recommandations du NCTM (1989, 1991, 1995).

Évaluation des systèmes d'éducation

Le Canada et ses provinces participent régulièrement à une série d'évaluations des apprentissages à grande échelle d'origines nationale et internationale, dans le but d'évaluer le fonctionnement de leurs systèmes d'éducation. La plupart de ces programmes d'évaluation comportent deux axes distincts, soit l'évaluation du rendement des élèves en une matière donnée et l'évaluation de multiples facteurs contextuels soupçonnés d'entretenir un lien avec le rendement des élèves. À cette fin, les programmes administrent des questionnaires contextuels aux élèves, aux parents, aux enseignants ainsi qu'aux directeurs d'école. Même si à maintes reprises, des études empiriques ont démontré l'importance des facteurs comme le milieu familial, le statut socio économique ou les caractéristiques internes des élèves sur leur rendement (Cheng et al., 2001; Clauser, 2001; Herry, 2000; Gales, 2000; Landry et al., 2002; Webster et al., 2000), récemment, les résultats et les données contextuelles portant sur le processus de fonctionnement des systèmes d'éducation sont davantage privilégiées.

Les croyances et les pratiques pédagogiques des enseignants comptent parmi ce type d'indicateurs. Plusieurs auteurs s'entendent pour dire que le développement d'indicateurs traitant des pratiques pédagogiques est très important pour améliorer la qualité de l'enseignement et de l'apprentissage ainsi que pour effectuer un suivi de la mise en œuvre des nouvelles politiques de réforme (Ball & Rowan, 2004; Clare, 2000; Linn & Baker, 1998; Masini & Taylor, 2000; Mullis, 2002; Scheerens, 2000; Thap & Gallimore, 1988). L'étude de O'Connor-Petruso, Schiering, Hayes et Serrano (2004), portant sur les facteurs pédagogiques et parentaux ayant un impact sur le rendement mathématiques des élèves de la Belgique, des Pays-bas et de la République Slovaque, les trois pays au sommet du palmarès européen de la TEIMS-Révisé de 1999, affirment que les facteurs pédagogiques sont déterminants pour le rendement des élèves. Plus précisément, les auteurs rapportent que les enseignants de ces pays privilégient des pratiques pédagogiques incitant au développement de la pensée critique et au développement des habiletés de résolution de problèmes.

En 2000, le North Central Regional Educational Laboratory (NCREL) aux États-Unis a également établi un lien entre certaines pratiques pédagogiques et le rendement des élèves en mathématiques. Il mentionne notamment que les exercices répétitifs corréleront négativement avec le rendement en mathématiques mais que les pratiques pédagogiques exigeant une compréhension approfondie des concepts mathématiques produisent des corrélations positives. En effet, les auteurs révèlent que les meilleurs enseignants mettent davantage l'accent sur la compréhension des concepts mathématiques que les autres enseignants des États-unis et qu'ils sont également plus susceptibles de demander à leurs élèves d'expliquer le raisonnement derrière une idée et d'écrire des équations pour représenter des relations. Ces enseignants accordent une plus grande importance au

raisonnement mathématique en comparaison à l'appropriation d'habiletés spécifiques et exigent que leurs élèves écrivent leurs idées, justifient leurs solutions et explorent diverses stratégies de résolution de problèmes (Masini & Taylor, 2000).

De tels résultats montrent à quel point les données générées par les programmes d'évaluation à grande échelle peuvent influencer l'élaboration de futures politiques éducatives. Cependant, peu d'études accompagnent la parution de leurs résultats par une mise en garde au sujet de la qualité des données initiales. Afin de pouvoir faire confiance aux conclusions de ces études et de s'assurer du bien-fondé des retombées politiques engendrées, des recherches additionnelles portant sur la qualité psychométrique des données provenant des évaluations à grande échelle s'imposent. L'examen de la qualité psychométrique des questionnaires contextuels comprend l'analyse de plusieurs concepts découlant des principes de validité et de fidélité. L'accumulation d'arguments ou de preuves dans le but d'établir le niveau de qualité s'articule autour de la validité des inférences formulées à partir des résultats issus d'un instrument donné (Bégin, 2006; Fraenkel & Wallen, 2003; Laveault & Grégoire, 2002).

Les données contextuelles issues des programmes d'évaluation à grande échelle proviennent en grande majorité de questionnaires administrés auprès des participants. Même si cet outil se veut relativement simple et peu coûteux et qu'il offre la possibilité d'amasser des quantités importantes d'information, les critiques se font nombreuses à son égard. Des auteurs ayant eu recours aux données des évaluations à grande échelle pour mener à bien des recherches sur les facteurs reliés à l'apprentissage des élèves rapportent que la qualité des données présente de sérieuses limites aux conclusions de leurs études (Anderson et al., 2004; Davidson, 1999; Gales, 2000; Hodges, 2000; Simon et al., 2005; Rodriguez, 1999;

Ungerleider, 2003). Selon Clare (2000), les questionnaires contextuels s'avèrent problématiques spécialement au niveau des questions portant sur les nouvelles politiques de réforme scolaire. Les auteurs ci-dessus rapportent qu'encore trop peu d'études ont été réalisées sur la qualité générale des résultats issus des questionnaires contextuels administrés aux enseignants et que la venue de nouveaux concepts introduits par les mouvements de réforme scolaire ont fort probablement des effets néfastes sur la qualité des données (Hill, 2002; Mayer 1999). Ils concluent que des recherches additionnelles devraient porter sur la qualité psychométrique des items qui constituent les questionnaires contextuels administrés aux enseignants et aux élèves, et plus précisément sur les items qui traitent des nouvelles pratiques pédagogiques instaurées par les récentes politiques éducatives.

Qualité psychométrique des questionnaires contextuels

Une recension des études traitant de la qualité des items de questionnaires contextuels destinés aux enseignants fût réalisée dans le cadre de la présente étude. Cette étape avait comme principal objectif de répertorier les différentes limites inhérentes aux données générées par ces instruments de mesure. La synthèse présentée dans cette section comporte les limites documentées au sujet d'items spécialement conçus pour évaluer les différentes pratiques et croyances pédagogiques des enseignants. Elle s'organise en fonction des concepts qui définissent le principe de qualité psychométrique, notamment : la validité de contenu, la validité liée à un critère, la validité théorique, la validité de conséquence, la validité apparente, le biais de réponse et la fidélité. Le tableau 2.1 répartit les études recensées et les limites rapportées selon les concepts de validité ou de fidélité. Les sections

subséquentes fournissent des détails au sujet des diverses limites répertoriées et des solutions envisagées par les auteurs.

Tableau 1

Limites rapportées au sujet des items traitant des pratiques pédagogiques

Concepts	Limites	Auteurs
Validité apparente et validité de contenu	- Mauvaise formulation de la question	Anderson, Rogers & Ungerleider (2004)
	- Absence de définitions pour les termes utilisés	Burstein (1995)
	- Présence et nombre de sous questions	Ball & Rowan (2004)
	- Problèmes liés à la période de référence de l'item (clarté et longueur appropriée pour le comportement ciblé)	Clare (2001)
	- Mauvaise représentativité des domaines à mesurer (Nombre insuffisant d'items)	Laveault & Grégoire (2002)
	- Manque de profondeur dans les questions	Mayer (1999)
	- Manque de discrimination ou de sensibilité des items	Mullens & Kasprzyk (1999)
	- Manque de efficacité des questionnaires pour mesurer des phénomènes complexes	NCES (1999, 2002)
	- Manque de pertinence du sujet de la question et des catégories de réponse proposées pour les enseignants devant y répondre et les chercheurs qui utilisent les données.	Porter & Smithson (2001)
		Rogers et al. (2004)
Validité liée à un critère	- Absence du processus de validation avec des critères externes pour les évaluations à grande échelle (opérations trop vastes, complexes et coûteuses).	Rowan, Correnti & Miller (2002)
		Burstein (1995)
		Mayer (1999)
		NCES (1999)
		Rowan, Correnti & Miller (2002)

Tableau 1 (suite)

Validité	- Absence de consensus pour les définitions des concepts	Burstein (1995)
théorique	<p>clés : les théoriciens ne s'entendent pas sur une définition claire de la « qualité d'enseignement » ni des « pratiques pédagogiques ».</p> <p>- Variation importante des politiques de réforme (Différence des pratiques pédagogiques proposées par les politiques et des définitions choisies).</p> <p>- Variation dans la façon dont les enseignants définissent les nouvelles pratiques pédagogiques</p> <p>- Manque de consensus au sujet de l'impact des nouvelles pratiques pédagogiques sur le rendement des élèves.</p>	<p>NCES (1999)</p> <p>Rowan, Correnti & Miller (2002)</p> <p>Spillane & Zeuli (1999)</p> <p>Rey et al. (2005)</p>
Validité de	- Diffusion de résultats non validés ou de piètre qualité	Anderson, Rogers & Ungerleider
conséquence	<p>- Prises de décisions non fondées - au sujet des nouvelles politiques de réforme souvent faites à partir de résultats non validés ou de piètre qualité</p> <p>- Production de variables presque uniquement ordinales ou dichotomiques</p> <p>- Analyses statistiques limitées en fonction du type de variables générées et de leurs caractéristiques de distribution (asymétrie, kurtose, variabilité, etc.)</p> <p>- Grandeur de l'effet des analyses corrélationnelles limitée par les variables catégorielles</p>	<p>(2004)</p> <p>Rogers, Anderson, Klinger & Dawber (2004)</p> <p>Mayer (1999)</p> <p>NCES (1999)</p>

Tableau 1 (suite)

Biais de réponse	- Présence d'un biais de réponse pour les enseignants	Burstein (1995)
	soucieux de répondre aux exigences des nouvelles	Mayer (1999)
	politiques de réforme (phénomène de désirabilité	Mullens & Kasprzyk (1999)
	sociale)	NCES (1999)
		Porter & Smithson (2001)
		Rowan, Correnti & Miller (2002)
		Spillane & Zeuli (1999)
Fidélité	- Petit nombre d'items dans les questionnaires	Ball & Rowan (2004)
	- Problèmes au niveau de la clarté de la question	Mayer (1999)
	Non-permanence des définitions de nouveaux mots issus	Mullens & Kasprzyk (1999)
	des réformes	NCES (1999, 2002)
	- Trop longue période de référence (difficultés de rappel)	Porter & Smithson (2001)
	- Faible coefficients de fidélité (cohérence interne) des indices composés créés à partir des variables traitant des pratiques pédagogiques du mouvement de réforme	Rowan, Correnti & Miller (2002)

Les premières limites rapportées dans le tableau résumé concernent les concepts de validité apparente et de contenu qui sont traités simultanément en raison du chevauchement de leurs critères. La démonstration de la validité apparente, consiste à mener une étude de surface de la validité des items du test ou du questionnaire (Laveault & Grégoire, 2002). Par ailleurs, la validité de contenu représente la capacité d'un instrument (ou de ses items) à mesurer et représenter toutes les facettes importantes d'un concept ciblé (Bégin, 2006). Ces types de validité portent autant sur le format de l'instrument que sur le contenu de ce dernier.

Plusieurs limites imposées par la conception des questionnaires contextuels pouvant affecter la validité apparente ou de contenu ont été répertoriées dans la recension des écrits.

D'abord, des essais réalisés sur les items des questionnaires du NAEP ont permis d'identifier des difficultés de compréhension chez les participants en fonction de la complexité de la formulation de la question en particulier lorsque la question comporte des mots nouveaux ou peu familiers, une occurrence particulièrement fréquente en contexte de réforme. En effet, certains experts (Ball & Rowan, 2004; Mullens & Kasprzyk, 1999; NCES, 1999, 2002; Porter & Smithson, 2001) rapportent que la validité s'améliore lorsque l'instrument de mesure définit clairement les termes nouveaux ou peu familiers. Puisque ceci ne se produit que très rarement dans les questionnaires des évaluations à grande échelle, les enseignants interprètent généralement le vocabulaire à leur façon. Pour pallier à cette problématique, plutôt que d'utiliser un vocabulaire élaboré, spécifique au contexte de réforme, Burstein (1995) suggèrent de formuler les questions en décrivant brièvement les pratiques afin d'uniformiser l'interprétation qu'en font les participants. Les auteurs de l'étude sur les items de la NAEP suggèrent également de recourir à des entrevues auprès des enseignants afin d'identifier d'éventuels problèmes d'interprétation liés à la formulation de la question (NCES, 1999; NCES, 2002).

La présence de sous-questions peut pareillement détériorer la validité de l'instrument. Pour alléger la tâche des répondants, les questionnaires contextuels des évaluations à grande échelle contiennent surtout des items dont l'énoncé principal s'applique à un grand nombre de terminaisons ou de sous-questions. Dans un de ses rapports au sujet de la qualité des items de questionnaires contextuels administrés aux élèves et aux enseignants du NAEP, le National Center for Education Statistics (NCES, 2002), aux Etats-Unis, affirme

que les répondants oublient le contexte de la question en présence d'un trop grand nombre de sous-questions. Dans le questionnaire de la TEIMS (2003), par exemple, un des énoncés demande aux participants de préciser si, au cours des deux dernières années, ils ont participé à six activités présentées en sous questions. Il se peut que les répondants ne considèrent plus la période de référence, soit « au cours des deux dernières années » lorsqu'ils répondent aux dernières sous-questions.

À ce sujet, des auteurs soulèvent que les items des questionnaires comportent souvent des problèmes au niveau de la période de référence (Mayer, 1999; Rowan, Correnti & Miller, 2002). L'appellation « période de référence » est utilisée pour désigner l'intervalle de temps agissant comme base de comparaison pour les comportements ciblés par les items d'un questionnaire (Pepermans, 1990). Les auteurs consultés stipulent notamment que la période de référence doit être clairement précisée pour chaque item. Pour éviter les problèmes de rappel, elle doit être courte, sans quoi, les enseignants ont tendance à fournir une estimation-devinette (NCES 1999, 2002). Par ailleurs, lorsque l'échelle réfère à des fréquences, la période doit être suffisamment longue pour que le comportement visé par la question puisse se reproduire plusieurs fois.

Toujours dans l'optique de la validité apparente ou de contenu, la représentativité du domaine ciblé prend également beaucoup d'importance. Cette étape exige que des experts formulent une appréciation de la représentation des facettes du domaine à mesurer dans le questionnaire. Cette étape dépend énormément de la précision avec laquelle le domaine à mesurer est lui-même défini (validité théorique). En ce qui concerne l'évaluation des pratiques pédagogiques, il ne suffit pas de s'assurer que les items couvrent l'ensemble des pratiques, mais il devient aussi nécessaire de pouvoir offrir de l'information en profondeur.

En effet, des auteurs (Mayer, 1999; Porter & Smithson, 2001; Stecher et al., 2002) ont récemment critiqué le recours aux questionnaires contextuels pour mesurer les pratiques pédagogiques puisque les items élaborés jusqu'à maintenant portent principalement sur leur fréquence d'utilisation sans fournir d'information sur « comment » ces pratiques sont appliquées. Ainsi, les données recueillies dans le cadre des évaluations à grande échelle ne savent pas distinguer entre un enseignant qui applique avec rigueur et maestria les nouvelles pratiques préconisées dans les politiques de réforme et un enseignant qui n'en fait qu'une tentative escamotée mais répétée.

Une autre limite rapportée par les auteurs concerne le manque de sensibilité des items (Burstein 1995). Spillane et Zeuli (1999) ont utilisé les items provenant de la TEIMS et d'autres sources pour classer les enseignants selon qu'ils appliquent peu, en partie, ou majoritairement, les pratiques de réforme. Selon ces auteurs, l'information recueillie actuellement par les items des questionnaires contextuels ne permet pas de distinguer de manière très précise entre ces niveaux. Les pratiques pédagogiques des enseignants évoluent rapidement dans un contexte de réforme et les items doivent être sensibles aux changements qui s'opèrent au niveau des croyances et des pratiques pédagogiques des enseignants.

De plus, en ce qui a trait aux évaluations à grande échelle, des entraves à la validité surviennent surtout lorsque le sujet de la question ou les catégories de réponse ne sont pas pertinents au contexte (Anderson, et al., 2004; Burstein, 1995; Mayer, 1999; NCES, 1999, 2002; Rogers et al., 2004). Par exemple, lorsqu'une question au sujet des pratiques pédagogiques enquête sur une activité de classe non usuelle dans certains pays à l'étude, l'item a tendance à afficher un faible taux de réponse. Ce faible taux de réponse illustre souvent la non-pertinence de l'item ou un manque de compréhension chez les enseignants.

En d'autres instances, ce sont les options de réponse à la question qui ne permettent pas aux enseignants de fournir une réplique fidèle à leur réalité.

La validité liée à un critère, la deuxième catégorie présentée au tableau 1, consiste à déterminer le degré de correspondance entre les résultats de l'instrument de mesure à l'étude et une seconde prise d'information servant de référence (Fraenkel & Wallen, 2003; Laveault & Grégoire, 2002). En ce qui a trait à la validation des questionnaires contextuels, ces mesures de comparaison comprennent par exemple, la collection d'artéfacts, l'observation en salle de classe et la prise de vidéo. Certaines études rapportent des indices de corrélation calculés entre le contenu d'un journal de bord complété par l'enseignant et ses réponses à un questionnaire contextuel (NCES, 1999; Porter & Smithson, 2001). Ceci s'avère inacceptable puisque ces deux instruments sont complétés par l'enseignant. Goldstein et al. proposent donc de comparer les réponses aux questionnaires des enseignants du National Assessment of Educational Progress (NAEP) à ceux de leurs élèves (NCES, 2002).

La troisième catégorie du tableau 1, la validité théorique, permet de donner un sens aux résultats obtenus. Les instruments de mesure sont conçus en fonction des phénomènes ou comportements *observables*, c'est-à-dire qu'ils reposent sur un modèle théorique de la réalité que l'on souhaite mesurer (Bégin, 2006; Laveault & Grégoire, 2002). Ces assises théoriques permettent d'interpréter les données empiriques recueillies et d'en dégager une signification. Selon Bégin (2006) la validité théorique d'un test sera d'autant possible à mettre en évidence que la «construction théorique» est solide et explicite. À ce sujet, la recherche menée par Coche, Kahn et Robin (2005) relate l'absence d'une définition claire du concept des pratiques pédagogiques. Il rapporte notamment les caractéristiques divergentes attribuées à la pratique pédagogique par différents chercheurs du domaine, dont M. Altet, M.

Durand, P. Perrenoud, B. Rey, M. Tardif et C. Lessard. Selon l'étude de Coche et al., les pratiques pédagogiques englobent parfois toutes les activités quotidiennes réalisées par l'enseignant dans le cadre de son travail, tandis qu'à d'autres moments, elles réfèrent uniquement aux activités délibérées et réfléchies d'enseignement. À d'autres moments, elles comprennent les activités d'apprentissage présentées aux élèves ou encore les différentes stratégies d'enseignement employées. Par exemple, est-ce que le fait de faire travailler les élèves en petits groupes représente une stratégie organisationnelle ou une pratique pédagogique avec un but d'apprentissage précis ? (Coche et al., 2005).

Les politiques de réforme varient également au niveau des pratiques pédagogiques proposées et de leurs définitions. Une brève comparaison des curriculums instaurés récemment dans deux provinces canadiennes laisse entendre qu'ils reposent sur des notions très similaires (p. ex., communication mathématique, raisonnement mathématique, résolution de problème, initiation aux technologies), mais qu'ils se distinguent surtout par le niveau de direction fourni aux enseignants en ce qui concerne les pratiques pédagogiques à préconisées.

Les opinions divergent également en ce qui concerne l'impact des nouvelles pratiques pédagogiques sur le rendement des élèves. En effet, plusieurs auteurs remettent en cause la qualité psychométrique des indicateurs pédagogiques, ce qui pourrait expliquer que le lien entre le rendement et l'enseignement soit difficile à démontrer (Burstein, 1995; NCES, 1999; Porter & Smithson, 2001; Rogers et al., 2004; Rowan, Correnti & Miller, 2002; Stecher & al., 2002). D'autres explications sont également fournies pour tenter de justifier cette lacune. D'abord les analyses corrélationnelles réalisées à partir de variables

catégorielles, souvent utilisées pour mesurer les croyances et les pratiques pédagogiques, affichent un effet statistique moindre.

La validité de conséquence, le quatrième concept du tableau 1, représente tous les impacts souhaités et non-souhaités de l'usage d'un instrument de mesure (Laveault & Grégoire, 2002). Les concepteurs des enquêtes et les utilisateurs des résultats de celles-ci doivent se soucier des conséquences directes et parfois indirectes engendrées par ces programmes. Puisque les résultats exercent une influence auprès des dirigeants chargés de prendre des décisions au sujet de systèmes d'éducation, leur manque de justesse peut mener à des interprétations erronées. Selon plusieurs chercheurs, le type de variable généré (dichotomique, catégorique, etc.) et les caractéristiques de distribution de ces variables limitent les analyses statistiques appropriées aux données des évaluations à grande échelle (Anderson et al., 2004; Rogers et al. 2004; Ungerleider, 2003). Le type de variable et leurs caractéristiques de distribution dépendent de la conception des items des questionnaires et de leur pertinence en contexte d'évaluation. Les conclusions produites à partir de ces analyses se voient influencées par la nature des analyses et la grandeur des effets statistiques.

Le sixième concept du tableau 1 porte sur la présence d'un biais de réponse. Ce concept sert à démontrer l'existence d'aucun biais systématique chez les répondants. Les concepteurs ou les utilisateurs de tests doivent accumuler des preuves dans le but de d'établir l'absence de biais envers un groupe de la population auquel le test sera appliqué (Laveault & Grégoire, 2002). En ce qui concerne les questionnaires contextuels à l'intention des enseignants, de nombreuses études dénoncent un biais résultant du phénomène de désirabilité sociale (Burstein , 1995; Laveault et Grégoire, 2002; Mayer, 1999; Mullens & Kasprzyk, 1999; NCES, 1999; Porter & Smithson, 2001; Rowan, Correnti & Miller, 2002;

Spillane & Zeuli, 1999). En effet, plusieurs de ces auteurs ont comparé les nouvelles pratiques pédagogiques rapportées dans un questionnaire complété par l'enseignant aux constats d'observations en salle de classe réalisés par des témoins. Les auteurs confirment que le phénomène de désirabilité sociale rejoint certains enseignants plus que d'autres en incitant ceux-ci à déclarer une utilisation plus fréquente de la pédagogie de réforme. Ce biais pourrait contribuer à fausser les résultats en ce qui concerne l'utilisation des nouvelles pratiques. De plus, les auteurs affirment qu'il est difficile de déterminer l'ordre de grandeur de cet effet.

Enfin, la fidélité d'un instrument de mesure, le septième concept du tableau 1, figure parmi les considérations les plus importantes et souvent les mieux documentées dans les enquêtes. Mayer (1999) rappelle que la fidélité dépend de la capacité d'un instrument de mesure à produire les mêmes résultats à chaque fois qu'il est administré. L'exercice le plus communément utilisé pour vérifier la fidélité d'un instrument est la méthode test- re-test qui a produit des résultats allant de bon à mauvais lors d'études précédentes effectuées sur les questionnaires contextuels administrés aux enseignants (Mayer, 1999; NCES, 1999; NCES, 2002). Les auteurs soulèvent diverses problématiques pouvant expliquer ces résultats comme la clarté de l'item, le manque de précision, la présence de mots nouveaux ou peu familiers. Un autre type de menace à la fidélité d'un item survient lorsque les périodes de référence sont trop longues ou mal précisées (Mayer 1999, Rowan, Correnti & Miller, 2002). En effet, une période de référence trop longue en termes de temps suscite parfois des difficultés de rappel (a déjà été discuté plus haut). Lorsque les concepteurs ou les utilisateurs de tests débattent de la fidélité d'un instrument, ils présentent également des résultats de cohérence interne à partir du coefficient *alpha* de Cronbach, une estimation de la proportion de

variance due au score vrai. Lorsque les items d'un test mesurent des concepts divergents, l'indice alpha présente un résultat faible. Parmi les études répertoriées, les résultats sont mitigés à ce sujet (Mayer, 1999). De façon générale, les items portant sur les pratiques pédagogiques regroupés en indices composés produisent des coefficients alpha moyens. Ceci signifie que ces items comportent une part significative d'erreur de mesure. Le coefficient alpha de Cronbach dépend du nombre d'items retenus dans la formation d'un indice composé.

En résumé, l'élaboration de nouvelles politiques éducatives appliquées dans les salles de classe est souvent inspirée par les résultats des évaluations à grande échelle. C'est pourquoi ces évaluations jouent un rôle crucial dans l'amélioration de la qualité de l'éducation dispensée dans les salles de classe (Mullis, 2002; Puk, 1999; Robitaille et Garden, 1996). Encore faut-il que ces données produites par les évaluations à grande échelle soient fiables. C'est pourquoi, il est nécessaire de mener des recherches additionnelles afin de déterminer la qualité psychométrique des questionnaires et de leurs produits, surtout lorsque ceux-ci portent sur la mise en œuvre des nouvelles politiques de réforme. C'est pourquoi la présente recherche se donne comme objectif d'évaluer la qualité psychométrique d'items issus des questionnaires contextuels des enseignants de la TEIMS portant sur les pratiques d'enseignement récemment introduites par les réformes scolaires ?

Même si certaines études de ce genre ont pu être répertoriées, elles présentent certaines lacunes importantes. D'abord, même si les études présentées dans la recension des écrits ciblent des objectifs de recherche analogues, d'une part elles ont recours à des méthodologies très variées pouvant mener à des difficultés d'interprétation et d'autre part les choix méthodologiques s'avèrent parfois plus ou moins justifiables. Pour les lecteurs, les

collègues chercheurs ou les utilisateurs de données, les conclusions présentées par de telles études doivent être clairement comprises et facilement interprétées. Bien qu'une grande diversité des moyens méthodologiques employés puisse fournir un regard plus global, il est parfois difficile pour les lecteurs d'évaluer les conclusions amenées par les chercheurs et la comparaison des résultats d'une étude à une autre devient très complexe. Puisque les items de questionnaire promettent d'expliquer certains facteurs liés au rendement des élèves et de renseigner les dirigeants au sujet de la mise en application des réformes, les choix méthodologiques des chercheurs rapportant le niveau de qualité psychométrique de ces items gagne en importance.

En précisant les besoins en matière d'information, les techniques méthodologiques appropriées pourront être perfectionnées et certaines lacunes repérées dans les études discutées dans la recension des écrits pourront être éliminées. Par exemple, lorsque les chercheurs du NCES (1999, 2002) comparent les résultats des enseignants à ceux de leurs élèves, ils ont choisi de réduire les échelles des items de quatre catégories de réponse à trois, afin d'augmenter les taux de concordance. Burstein (1995) opère de façon semblable lorsqu'il rapporte les taux de concordance pour des items de quatre ou de cinq catégories de réponse. Afin d'augmenter le taux d'accord entre les élèves et les enseignants, il choisit de rapporter les taux en fonction des concordances « within one » dans lequel cas, une réponse de 3 serait considérée comme étant équivalente à une réponse de 2, 3, ou 4. Cela a pour effet d'augmenter de façon considérable les résultats de concordance. D'autres anomalies méthodologiques surviennent notamment lorsque les résultats des questionnaires sont comparés à ceux des journaux de bord également complétés par les enseignants. D'abord, Mayer (1999) affirme qu'une validation de ce genre ne porte aucune valeur puisqu'elle

devrait être effectuée uniquement entre deux sources d'information différentes, i.e. que les journaux de bord ne pourraient pas être utilisés pour valider les questionnaires des enseignants puisque les deux instruments recueillent des renseignements auprès des mêmes participants. De plus, lors de la comparaison des résultats produits par les journaux de bord et les questionnaires (NCES, 1999), les options de réponse et les périodes de référence des questionnaires ne correspondaient pas parfaitement à ceux des journaux du bord.

D'autre part, plusieurs des recherches répertoriées dans le cadre de cette étude présentent des méthodologies principalement quantitatives. Bien que les procédés statistiques détiennent le potentiel de fournir une quantité importante d'information au sujet de la qualité psychométrique des items de questionnaires, les analyses qualitatives s'avèrent tout aussi riches et prometteuses dans ce contexte. En effet, si les analyses quantitatives peuvent servir dans un premier temps à identifier les items problématiques, le recours aux méthodes qualitatives permet d'expliquer pourquoi certains items ne produisent pas les résultats escomptés. Ces deux choix méthodologiques sont complémentaires et les études portant sur la qualité des items de questionnaire ont intérêt à intégrer ces deux formes d'analyse dans le but de produire des résultats davantage riches, significatifs et concluants.

Pour terminer, la majorité des études consultées dans le cadre de ce projet ont été effectuées pendant la phase de sélection ou de raffinement des items composant une version d'essai ou finale d'un questionnaire. Certaines de ces études, semblaient vouloir faire valoir la bonne qualité des items composant le questionnaire conçu et cela a pu biaiser le choix de la méthodologie et la présentation des résultats. Idéalement, ce type d'étude devrait se réaliser au début d'un processus de construction de test ou lors de la mise à l'essai d'items. En outre, pour éviter que les résultats de ces études soient biaisés, elles devraient être

menées par un parti neutre et indépendant, c'est-à-dire des chercheurs n'étant pas responsable de la création des items ou de l'administration des questionnaires. Entre autres, la recension a permis d'identifier des études ciblant la qualité des items du NAEP, toutefois, peu d'études du genre portent sur d'autres programmes d'évaluation à grande échelle, surtout lorsqu'il s'agit de programmes internationaux tels que le TEIMS, le Programme international de recherche en lecture scolaire (PIRLS) ou le Programme international pour le suivi des acquis des élèves (PISA).

En conclusion, la présente recherche se donne comme objectif d'évaluer la qualité psychométrique d'items issus du questionnaire contextuel des enseignants élaboré par la TEIMS dans le cadre des cycles d'enquête de 1995, 1999 et de 2003. Les items retenus pour l'étude portent sur les pratiques pédagogiques ciblées par les récents mouvements de réforme scolaire dans le domaine des mathématiques. Ainsi, les résultats de l'étude aideront à déterminer si la qualité des données contextuelles issues des programmes d'évaluation à grande échelle permet une utilisation par les dirigeants en vue de leur planification des curriculums et de l'allocation des ressources.

L'étude se donne comme objectif principal de répondre à la question de recherche suivante : Quelle est la qualité psychométrique des items issus des questionnaires contextuels des enseignants de la TEIMS portant sur les pratiques d'enseignement (curriculum appliqué) ciblées par les réformes scolaires (curriculum projeté) ? La section suivante présente les éléments du cadre de référence permettant d'examiner les qualités psychométriques pertinentes en vue de répondre à la question de recherche.

Cadre de référence

Le cadre de référence retenu pour la présente étude s'inspire des fondements conceptuels sous-jacents au développement de la TEIMS (Robitaille et Garden, 1996; Travers & Westbury, 1989) et aux qualités psychométriques des questionnaires contextuels en fonction des construits ciblés dans les réformes scolaires. Ces deux éléments structurent la présente section.

Cadre sous-jacent à la TEIMS

Le cadre sous-jacent à la TEIMS comprend principalement trois grandes thématiques : le curriculum projeté, le curriculum appliqué et le curriculum atteint. Il s'agit en effet du schème de référence utilisé pour la conception des questionnaires contextuels destinés aux écoles et aux enseignants et certaines composantes des questionnaires des élèves. Dans ce cadre, le curriculum projeté se veut une déclaration d'une société quant à ses objectifs d'enseignement et d'apprentissage. Il se traduit en manuels scolaires, guides et curriculums, évaluations du rendement et politiques éducatives (Robitaille & Garden, 1996, p.35). La réforme scolaire se définit comme « un changement majeur, souhaitable et valide d'une situation, d'une chose ou d'une institution, d'un état initial à un état projeté » (Legendre, 1993, p. 1100). Legendre précise également que le terme « réforme », dans un contexte éducatif, s'inscrit nécessairement dans l'optique d'un changement *en bien*, contrairement au concept de l'évolution.

Par ailleurs, le curriculum appliqué se réfère aux pratiques pédagogiques, c'est-à-dire à la nature, au contenu et aux moyens adoptés lors de la mise en oeuvre du curriculum

projeté. L'administration de l'école et du conseil scolaire, la communauté, les ressources disponibles et les caractéristiques de la classe influencent le curriculum appliqué (Robitaille & Garden, 1996, p.36). Les questionnaires contextuels de la TEIMS à l'intention des enseignants portent sur les aspects suivants du curriculum appliqué : les caractéristiques démographiques comme l'âge, le niveau d'éducation, les années d'expérience, les croyances pédagogiques et le contexte de la salle de classe où on retrouve l'accès aux ressources éducatives, les pratiques pédagogiques, l'organisation de la classe, le matériel pédagogique, les stratégies et les activités pédagogiques, les devoirs, l'évaluation et le contenu de la matière enseignée (Mullis, 2002; Schmidt & Cogan, 1996). De plus, les questionnaires contextuels des élèves reprennent de façon quasi-identique plusieurs des items portant sur les pratiques pédagogiques de leur enseignant.

Enfin, le curriculum atteint concerne la partie du curriculum projeté qui est assimilée par les élèves et qui se manifeste dans leur rendement et leurs attitudes face à une matière donnée (Robitaille & Garden, 1996, p.35). La figure 1 illustre le cadre de référence proposé. Elle démontre l'interdépendance des trois niveaux de curriculum, mais uniquement à partir du contexte le plus large vers le plus précis. En effet, le curriculum projeté sert d'assises aux deux autres niveaux. Le curriculum appliqué, dépend inévitablement du curriculum projeté qui sert de point de départ mais qui se voit également influencé par plusieurs autres facteurs. L'influence peut venir de facteurs situés au niveau de l'individu (enseignant), où notamment le niveau de connaissance des concepts mathématiques ou les pratiques pédagogiques privilégiées par l'enseignant jouent un rôle important. L'influence peut également provenir du milieu dans lequel l'individu se trouve, en supposant, par exemple, que la philosophie véhiculée par un conseil scolaire puisse guider l'application du curriculum par les

enseignants. Bref, c'est donc l'interprétation du curriculum projeté par les acteurs du curriculum appliqué qui viendra agir sur le curriculum atteint.

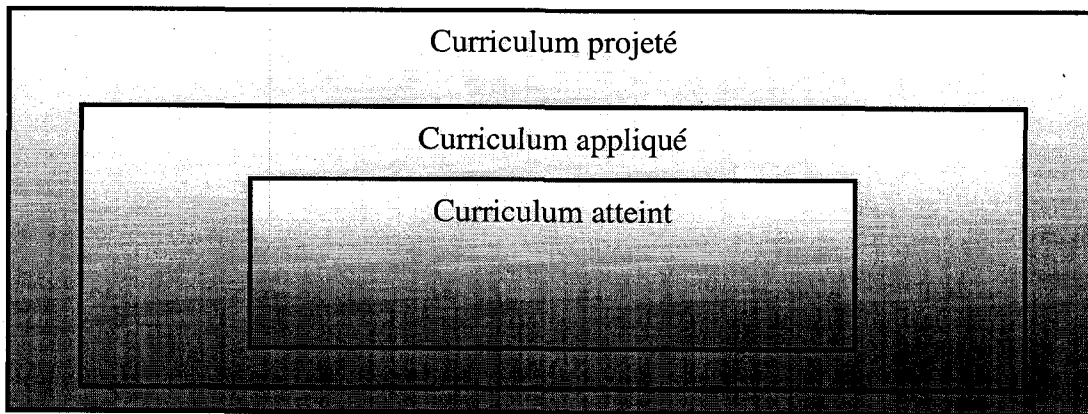


Figure 1. Interactions entre les trois niveaux de curriculum.

La présente étude tient compte des trois niveaux du curriculum mais traite des données des questionnaires contextuels uniquement. Elle vise à déterminer la qualité des moyens privilégiés par les programmes d'évaluation à grande échelle en vue d'obtenir une mesure du curriculum appliqué en contexte de réforme scolaire. Puisque l'échantillon est sélectionné dans la province de l'Ontario, son nouveau curriculum de l'élémentaire (Ministère de l'Éducation, 1997) représente le curriculum projeté de cette étude. Le curriculum appliqué dans son sens large comprend à la fois le contenu enseigné et la façon dont ce contenu est présenté en salle de classe. Puisque d'autres études ont déjà fait l'examen des contenus enseignés en Ontario (Lawson, Bordignon & Nagy, 2002) et que l'importance des pratiques pédagogiques prend de l'ampleur dans les récentes réformes scolaires, la présente étude se concentre uniquement sur les pratiques des enseignants.

L'étude se donne comme objectif principal de répondre à la question de recherche suivante : Quelle est la qualité psychométrique des items issus des questionnaires contextuels des enseignants de la TEIMS portant sur les pratiques d'enseignement (curriculum appliqué) ciblées par les réformes scolaires (curriculum projeté) ? La section suivante présente les éléments du cadre de référence permettant d'examiner les qualités psychométriques pertinentes en vue de répondre à la question de recherche.

Cadre sous-jacent à la qualité psychométrique du questionnaire contextuel

L'obtention d'une réponse claire et utile à la question de recherche nécessite la planification d'une série d'analyses quantitatives et qualitatives organisées en fonction d'aspects clés de la qualité psychométrique. Le tableau 2 présente cinq concepts retenus en vue de déterminer le niveau de la qualité psychométrique des items du questionnaire contextuel qui portent sur les pratiques pédagogiques récentes. Il reprend essentiellement les concepts reliés à la validité et à la fidélité présentés dans la section de la recension des écrits. Le tableau exclut toutefois certains concepts dont il serait difficile de mesurer sans obtenir, post-hoc, la participation des sujets initiaux de la TEIMS.

Tableau 2

Éléments constitutifs du cadre psychométrique

Concept à mesurer	Définition du concept	Critères évalués
<i>Validité</i>		
Validité apparente et Validité de contenu	Capacité des questionnaires contextuels à mesurer les principales pratiques pédagogiques suggérées par la politique de réforme.	<ul style="list-style-type: none"> ▪ Formulation de la question ▪ Clarté de la question ▪ Présence de définition pour le vocabulaire utilisé ▪ Niveau de familiarité du vocabulaire utilisé ▪ Présence de plus d'un concept dans la question ▪ Pertinence du sujet abordé par l'item pour le contexte éducatif ▪ Pertinence des choix de réponse ▪ Présence et nombre de sous-questions ▪ Période de référence de l'item ▪ Sensibilité des items
Concordance élève/enseignant	Degré de concordance entre les réponses fournies par les enseignants et celles des observateurs (élèves) au sujet des pratiques pédagogiques employées en salle de classe.	<ul style="list-style-type: none"> ▪ Évaluation du degré de déviation moyenne entre les résultats des questionnaires contextuels enseignant et élève

Tableau 2 (suite)

Validité théorique	Établissement du construit en fonction des trois compétences pédagogiques visées par le curriculum	▪ Cohérence des modèles exploratoires issus des analyses factorielles
Biais de réponse	Évaluation de la présence d'un biais de réponse	▪ Présence d'un biais de réponse rapporté par des enseignants, dû au phénomène de désirabilité sociale
<i>Fidélité</i>		
Fidélité	Examen des indices composés	▪ Méthode des covariances (cohérence interne)

Dans le contexte de la présente étude, la validité apparente sert à décrire le niveau de validité de surface pressenti au sujet des items sélectionnés du questionnaire des enseignants. La validité de contenu représente quant à elle, la capacité qu'ont les questionnaires contextuels de la TEIMS à mesurer, chez les enseignants sondés, la prévalence des pratiques et des croyances pédagogiques du curriculum de 1997. L'évaluation de ces concepts requiert un examen du contenu et du format des items du questionnaire. Précisément, une série d'éléments servent à dresser un portrait de ces formes de validité, soit : la formulation de la question, la clarté de la question, la présence de définition du vocabulaire utilisé, le niveau de familiarité du vocabulaire, la pertinence du sujet abordé par l'item et ses choix de réponse en fonction du contexte éducatif, la présence et le nombre de sous-questions, la présence de

plus d'un concept dans la question, la période de référence de l'item et la sensibilité des items.

En ce qui a trait à la concordance entre élève et enseignant, elle se traduit ici par le rapport qui existe entre les résultats des items des questionnaires enseignants portant sur les nouvelles pratiques pédagogiques et une autre source de données. Cette validation s'effectue en comparant l'information rapportée par les enseignants à celle compilée à partir des propos de leurs élèves. La recension des écrits a déjà fait état des propos de certains auteurs au sujet des problèmes engendrés par les réponses biaisées des enseignants rapportant leurs propres pratiques (Mayer, 1999; Mullens & Kasprzyk, 1999; NCES, 1999; Porter & Smithson, 2001; Rowan, Correnti & Miller, 2002). Cette opération permettra de valider l'information rapportée par les enseignants au sujet des pratiques préconisées en salle de classe.

Pour ce qui est de la validité théorique, elle est établie dans le but de déterminer le sens devant être accordé aux résultats d'une étude. Tout instrument de mesure repose sur un concept ou modèle théorique d'une réalité que l'on tente de mesurer. Ce modèle, implicite ou explicite, permet d'interpréter les données recueillies et de leur donner un sens. On doit non seulement analyser les questionnaires mais également réaliser une étude approfondie du curriculum afin d'établir la validité théorique. Pour la présente étude, l'analyse factorielle vise à vérifier si les réponses des enseignants sont le reflet des concepts théoriques à la base de la conception du questionnaire. La validation du construit s'élabore selon les trois compétences pédagogiques visées : raisonnement /compréhension mathématique, résolution de problème, et communication mathématique. Elle aide à déterminer dans quelle mesure les items retenus mesurent réellement des concepts pédagogiques sous-jacents communs.

Le concept suivant requiert une analyse pour déterminer s'il y a présence d'un biais de réponse chez les participants de l'étude. Pour la situation actuelle, seul le biais de réponse associé au concept de « désirabilité sociale » sera investigué. Pour se faire, des questions sont adressées aux participants qui doivent fournir leurs impressions au sujet de la présence d'un tel biais de réponse et d'un sens moral ou professionnel associé aux items des questionnaires contextuels de la TEIMS.

Finalement, la fidélité d'un instrument représente la capacité d'un test à reproduire les mêmes résultats dans les mêmes conditions. La théorie classique des tests stipule que chaque mesure comporte une part de score vrai et une part d'erreur de mesure. Une mesure fidèle arrive à cibler majoritairement le score vrai et donc comporte peu d'erreur de mesure. Le concept de fidélité d'un item est quantifié par la proportion de variabilité du score observé qui est due à la variabilité du score vrai (StatSoft, 2006). La théorie classique des tests contient également de l'information quant à l'amalgame de plusieurs items pour représenter différents concepts. L'erreur étant mesurée par tous les items est estimée être totalement aléatoire, cela permet de démontrer que la moyenne des ces termes d'erreur s'approche de zéro. De cette façon, plus il y a d'items mesurant un même concept, plus la proportion de score vrai devrait être importante comparativement à la portion d'erreur. Cette logique s'avère vraie uniquement si les items entre-eux mesurent réellement un concept commun et si bien-sûr un autre biais systématique ne vient pas annuler la supposition que l'erreur soit aléatoire. Dans le cadre de cette étude, seul la fidélité des indices composés sera ciblée puisque les autres démarches exigent une seconde prise d'information auprès des participants initiaux de la TEIMS.

L'examen de ces concepts permet de mettre en évidence les qualités et les limites inhérentes aux items des questionnaires contextuels administrés aux enseignants. La portée de cette étude rejoint uniquement les items traitant des nouvelles pratiques pédagogiques instaurées lors des récents courants de renouveau pédagogique. Les pédagogies incluses dans le spectre de la présente étude, font l'unanimité entre les chercheurs et réformateurs du domaine et figurent parmi les objectifs de nombreuses politiques de réforme aux États-unis et au Canada (Burstein, 1995; Clare, 2000; Masini & Taylor, 2000; Mayer, 1999; NCES, 1999; Porter & Smithson, 2001; Spillane & Zeuli, 1999; Stecher et al., 2002). Le curriculum de l'Ontario réserve une place de choix à ces mêmes pédagogies puisqu'elles représentent trois des quatre compétences centrales visées par le curriculum. En effet, le nouveau curriculum de la province de l'Ontario s'organise selon quatre compétences, et chacune d'entre-elles correspond aux recommandations du NCTM, elles sont : la résolution de problèmes, la compréhension des concepts, l'application des procédures mathématiques et la communication (MEFO, 1997). Le curriculum prescrit notamment aux enseignants de structurer leur enseignement de façon à ce que la résolution de problèmes complexes devienne une pierre angulaire lors de la transmission des notions mathématiques aux élèves. Il requiert également que les élèves justifient leur raisonnement et qu'ils expliquent leurs réponses. Les activités d'apprentissage doivent permettre aux élèves de travailler en équipes et donc de communiquer leur processus mathématique avec leurs pairs et leur enseignant. De plus, le nouveau curriculum sollicite l'application concrète et pratique des concepts et des théories mathématiques dans la vie quotidienne des élèves. Selon le MEFO, les mathématiques doivent être considérées comme un ensemble de règles qui couvrent toutes les possibilités et l'exploration des concepts mathématiques, la découverte et l'application de

règles devraient se faire par la manipulation d'objets divers ainsi que par l'utilisation des ordinateurs, de calculatrices et d'autres technologies. On y suggère également d'utiliser une grande variété de pratiques pédagogiques afin de rencontrer tous les objectifs fixés pour l'apprentissage des élèves (MEFO, 1997).

Les trois catégories de pratiques retenues pour cette étude sont plus précisément le raisonnement/compréhension mathématique, la résolution de problèmes et la communication mathématique. La section qui suit comprend des extraits tirés du curriculum de l'Ontario de 1997 décrivant les trois compétences visées par la présente étude. Ces extraits servent à définir de façon plus explicite ces compétences pédagogiques tel qu'elles sont véhiculées par le MEFO .

D'abord, les enseignants devraient mettre l'accent sur le raisonnement mathématique afin que leurs élèves puissent mieux comprendre le monde qui les entoure – i.e. des liens directs doivent être établis entre les mathématiques et la vie quotidienne des élèves. Le Ministère de l'éducation (1997) propose aussi que pour développer la compétence du raisonnement mathématique, les enseignants devraient créer un climat de classe axé sur la pensée critique ce qui requiert une approche systématique, analytique et bien raisonnée de l'apprentissage des concepts et des processus mathématiques (p. 68). Il suggère notamment de stimuler l'esprit critique et investigateur des élèves en ayant fréquemment recours à des questions comme « Pourquoi penses-tu que c'est une bonne réponse? » ou « Penses-tu que tu aurais obtenu la même réponse si tu avais utilisé ces autres matériaux? » (p. 68). Le raisonnement et la compréhension des concepts mathématiques sont encouragés par l'établissement de liens, ainsi que l'étude des régularités et des relations (p. 46). Le curriculum propose également de faire appel à du matériel concret et à des modèles pour

comprendre les concepts et expliquer le raisonnement (p. 69). En outre, l'utilisation des nouvelles technologies comme les calculatrices et les ordinateurs joue un rôle essentiel car selon le MEFO, elle facilite l'accès au raisonnement abstrait et permet une exploration créatrice des concepts mathématiques (p.70).

Par ailleurs, le curriculum de l'Ontario stipule que le développement de la compétence de résolution de problèmes mathématiques fait appel à des activités qui dépassent les exercices répétitifs, et que cela contribue à soutenir l'intérêt des élèves et leur motivation pour l'apprentissage (p. 6). Il ne s'agit pas de demander aux élèves de pratiquer leurs aptitudes de résolution de problèmes uniquement pour répondre à des énoncés sous forme de texte. Dans le nouveau curriculum de l'Ontario, il est plutôt question d'organiser l'enseignement et l'apprentissage de façon à ce que la résolution de problèmes occupe un rôle central (p. 65). L'élève devrait avoir la chance d'exercer les attitudes, stratégies et méthodes de réflexion dans des tâches de résolution complexes et significatives lui permettant également de faire des liens entre le monde réel et les mathématiques (p.65). Les nouvelles technologies (ordinateurs et calculatrices) représentent des outils importants de résolution de problèmes pour les élèves, ils doivent apprendre à les utiliser efficacement pour analyser des problèmes et présenter leurs solutions (p. 5). La gamme complète des habiletés et des connaissances mathématiques peut être enseignée à travers la résolution de problèmes surtout si ces derniers comportent des éléments provenant des cinq domaines d'études et présentent divers contextes (p. 66).

Enfin, le curriculum propose des pratiques d'enseignement favorisant la communication mathématique. Il décrit notamment les mathématiques comme un outil de communication précis et efficace pouvant servir même aux plus jeunes élèves à exprimer

leurs idées avec clarté, à mieux raisonner et à justifier leurs conclusions. Des questions comme « Pourquoi ? », « Que se passera-t-il si...? », « Peux-tu me donner un exemple de... ? », « Y a-t-il un exemple du contraire ? », « Vois-tu une régularité ? », « Est-ce toujours vrai? » peuvent aider les élèves à approfondir leur capacité d'évaluer et de défendre leurs idées (p. 70). La connaissance du langage, des structures et des opérations mathématiques soutient l'apprentissage des mathématiques. Le MEFO affirme également que la communication aide les élèves à mieux résoudre des problèmes mathématiques et recommande que celle-ci soit intégrée à toutes les étapes du processus de résolution (p. 67). Et finalement, un climat de respect mutuel et de soutien dans la classe et le recours aux travaux de groupes permettra aux élèves de s'exprimer librement, de questionner et de réagir aux idées avancées par leurs pairs ainsi que par l'enseignante (p. 68).

En résumé, dans le cadre de cette étude, des items du questionnaire contextuel à l'intention des enseignants feront l'objet d'analyses qualitatives et quantitatives, selon les concepts de validité apparente, de contenu, et théorique, de concordance élève/enseignant, de biais de réponse ainsi que de fidélité. Les items ainsi évalués rejoignent les trois compétences suivantes : le raisonnement/compréhension mathématique, la résolution de problèmes et la communication mathématique. En ce sens, le Ministère de l'éducation et de la formation de l'Ontario (MEFO) propose aux enseignants de l'élémentaire un éventail de pratiques pédagogiques afin de promouvoir l'atteinte de ces compétences centrales. Inspiré du NCTM, le MEFO suggère notamment de miser sur des pratiques pédagogiques telles que, l'utilisation des nouvelles technologies, la justification explicite du raisonnement, l'ancrage des problèmes mathématiques dans la vie quotidienne de l'élève, le travail de groupes et une conception des mathématiques sous forme de règles et de régularités. La sélection des items

est basée sur les descriptions de ces concepts pédagogiques fournies par le curriculum de l'Ontario puisque l'étude cible la capacité du questionnaire à capter les pratiques pédagogiques nouvellement instaurées. Ces items seront ensuite évalués en fonction du principe de la qualité psychométrique.

Méthodologie

La présente étude comporte une méthodologie mixte puisqu'elle comporte deux composantes de recherche, une qui fasse appel à des données quantitatives et une deuxième qui soit qualitative (Johnson & Christensen, 2004). La section qui suit décrit les participants de l'étude, les instruments de mesure utilisés pour la collecte des données, les procédures de l'étude ainsi que le traitement des données réalisé.

Participants

Deux populations distinctes font partie de l'étude. La première compte tous les enseignants de mathématiques des écoles de langue française de 7^e et 8^e année de la province de l'Ontario et la deuxième comprend les élèves de 7^e et de 8^e année provenant de ces mêmes écoles. Aux fins de l'étude, seulement les enseignants qui enseignent les mathématiques dans les écoles de langue française ont été retenus puisque l'application du curriculum du ministère provincial et le matériel pédagogique varient entre les écoles de langue française et anglaise de l'Ontario. Cela, s'avère vrai même si le programme cadre est sensiblement le même pour ces deux groupes linguistiques.

Afin de représenter cette première population, un échantillon d'enseignants a été retenu pour chacun des trois cycles d'enquête de la TEIMS-R; soit, 1995, 1999 et 2003. Les enseignants faisant partie de ces trois échantillons, ont tout d'abord été sélectionnés par le processus d'échantillonnage de la TEIMS. La technique d'échantillonnage sélectionnée par la TEIMS se veut une méthode aléatoire à deux étapes qui cible principalement la sélection des élèves selon des classes entières, et par conséquent les enseignants associés à ces classes font également partie de l'échantillon. La première étape vise à identifier les écoles selon des strates données comme la situation géographique et le nombre d'élèves par école. Ensuite,

une méthode aléatoire par grappes permet de choisir une classe ou des classes dans chaque école. Au sein des classes sélectionnées, l'enseignant et ses élèves respectifs participent à l'enquête. Le plan d'échantillonnage de la TEIMS exclut les écoles en région très éloignée géographiquement, les écoles de très petite taille et les écoles qui desservent exclusivement une population d'élèves déficients ou ayant des difficultés d'apprentissages ne permettant pas la passation de la TEIMS (Foy & Joncas, 2001). Le tableau 3 contient l'information au sujet des échantillons de la présente étude. Tous les enseignants pour qui les dossiers étaient disponibles dans les fichiers de données de l'Ontario ont été considérés pour la création des échantillons de cette étude, et ce, peu importe leur niveau d'éducation, leur âge, le nombre d'années d'enseignement ou autres caractéristiques personnelles. Exceptionnellement, lorsque le dossier d'un participant contenait plus de 30 % de données manquantes pour les items ciblés par l'étude, il n'a pas été retenu dans l'échantillon final. Les échantillons finaux contiennent 123 participants en 1995, 90 pour 1999 et 75 pour 2003. La taille des échantillons représente une proportion importante de la population des enseignants de mathématiques francophones oeuvrant en Ontario en 7e ou 8e années.

Afin de pouvoir comparer les résultats des enseignants à ceux de leurs élèves pour obtenir une mesure de validité liée à un critère, la deuxième population ciblée par l'étude comprend tous les élèves de 7e et de 8e année étant inscrits dans une école française de l'Ontario et cheminant dans des classes régulières (Population 2 –TEIMS). Les trois échantillons d'élèves qui ont participé respectivement à la TEIMS de 1995, de 1999 et de 2003 font partie de l'étude. Les enseignants ont d'abord été sélectionnés selon les critères mentionnés ci-dessus et ensuite tous les dossiers de leurs étudiants furent ajoutés à l'échantillon final. Les analyses planifiées sur ces données permettent de conserver même

les enregistrements des élèves qui contiennent certaines données manquantes. Pour cette raison, ce sont uniquement les dossiers des élèves dont la totalité des items ciblés par la présente étude étaient absents au dossier qui ont été supprimés.

Tableau 3

Nombre d'enseignants, de classes et d'élèves dans les trois échantillons

	1995	1999	2003	Total
Participants				
Élèves	2409	2116	1473	5998
Classes	123	90	77	290
Enseignants	123	90	75	288

Les données obtenues à partir des échantillons décrits précédemment ont servi à mener l'ensemble des analyses statistiques quantitatives. Cependant, l'étude comporte aussi un aspect qualitatif, exigeant la collecte d'un ensemble de données additionnelles. Ces données qualitatives ont été recueillies auprès d'un panel de quatre enseignants de mathématiques des écoles de langue française de l'Est ontarien. Ces participants ont volontairement et généreusement offert de fournir les renseignements demandés au sujet des items de la TEIMS.

Instruments

Deux types d'instruments ont servi à la collecte de données aux fins de l'étude : le questionnaire contextuel de la TEIMS à l'intention des enseignants et un questionnaire à l'intention d'un panel d'enseignants élaboré dans le cadre de la présente étude.

Questionnaires contextuels de la TEIMS.

Étant de grande envergure, la TEIMS comporte plusieurs instruments de mesure destinés aux participants : test de rendement des élèves, questionnaire contextuel des élèves, questionnaire contextuel des enseignants et questionnaire contextuel des écoles. Pour répondre à la question de recherche, l'étude se concentre uniquement sur les questionnaires contextuels de l'enseignant et de l'élève, généralement conçus pour un temps d'administration de 30 minutes. Les items retenus dans les questionnaires contextuels de l'enseignant et de l'élève comportent des variables dichotomiques, catégorielles ordinales ou à intervalles affichant de trois à cinq catégories de réponse. La plupart de ces items évaluent la fréquence ou l'importance des diverses pratiques et croyances pédagogiques adoptées par les enseignants. Par exemple, l'item « *Pendant les cours de mathématiques, vous arrive-t-il de demander aux élèves de réaliser les activités suivantes : expliquer le raisonnement qui soutend une idée ?* » offre les options de réponse suivantes : jamais ou presque jamais, parfois, la plupart du temps, à chaque fois.

Les items choisis pour la présente étude à partir des questionnaires contextuels de la TEIMS portent plus spécifiquement sur les pratiques pédagogiques recommandées par le nouveau curriculum de l'élémentaire en Ontario (Ministère de l'Éducation et de la

Formation de l'Ontario, 1997). Entre les cycles de l'étude TEIMS (1995 – 1999 – 2003), des concertations ont lieu afin d'ajuster les instruments de mesure mais puisque toute modification limite le potentiel de comparaison entre les cycles, celles-ci sont généralement gardées à un minimum. Dans le cadre du présent projet de recherche, cela signifie que le questionnaire contextuel des enseignants de 1995 ne correspond pas de façon identique au questionnaire de 1999, qui à son tour ne correspond pas entièrement au questionnaire de 2003. Le contenu général des questionnaires se fonde toujours sur le même cadre conceptuel et les mêmes thèmes sont abordés, mais la formulation des questions et certains contenus varient légèrement d'un cycle à l'autre. La sélection des items s'est opérée en fonction des trois thèmes suivants :

- 1) Le raisonnement/compréhension mathématique
- 2) La résolution de problème
- 3) La communication mathématique

Les items des questionnaires enseignants de la TEIMS ont été évalués un à un en fonction de leur contenu. La chercheuse a utilisé des critères précis lors de la sélection des items, i.e. les items retenus répondent à un des deux critères suivants : 1) leur contenu est directement relié à une des trois compétences, ou 2) leur contenu rejoint directement une des pratiques pédagogiques proposées par le MEFO pour l'atteinte des trois compétences. La sélection finale des items présentée au tableau 5, fut réalisée en fonction de l'information explicitement énoncée par le curriculum de l'Ontario (1997) au sujet des compétences et des pratiques pédagogiques à préconiser pour l'enseignement des mathématiques. Toutefois, cette étape de sélection et de classification n'a pas été répétée avec d'autres experts, et donc, aucune validation de la sélection finale des items n'a été effectuée.

De plus, dans le but de réaliser certaines des analyses statistiques précises, les données issues des questionnaires contextuels à l'intention des élèves de la TEIMS ont dû être utilisées. Certaines questions demandées aux enseignants au sujet de leurs pratiques pédagogiques ont aussi été administrées aux élèves de leur classe. Ces items ont été sélectionnés pour les analyses. Le tableau 4 contient le nombre d'items retenus pour les questionnaires de l'enseignant et de l'élève en fonction des cycles de la TEIMS. Par exemple, pour 1995, 23 items du questionnaire enseignant ont été sélectionnés, parmi ces derniers, 20 rejoignent le concept du raisonnement, 11 celui de la résolution de problèmes et huit touchent à la communication mathématique. Il est à noter que certains items sont identifiés à plus d'un concept pédagogique. Toujours pour les données de 1995, cinq items du questionnaire élève ont été retenus, dont trois qui sont liés au raisonnement, deux à la résolution de problèmes et un à la communication.

Tableau 4

Nombre d'items retenus à partir des questionnaires enseignants et élèves en fonction des compétences pédagogiques ciblées

Compétences pédagogiques	Échantillons			
	1995 et 1999		2003	
	Enseignants	Élèves	Enseignants	Élèves
Raisonnement/ compréhension	20	3	19	7
Résolution de problèmes	11	2	15	5
Communication	8	1	7	5
Total des items retenus	23	5	26	10

Lors de la sélection des items, une comparaison entre les cycles a également été réalisée. Cette comparaison fût effectuée dans le but de déterminer quels items étaient repris façon consistante d'un cycle à un autre dans les questionnaires des enseignants et des élèves. L'inventaire réalisé paraît dans un tableau présenté en annexe B, où la disponibilité des items pour chacun des différents questionnaires TEIMS est indiquée par un X.

Questionnaire à l'intention du panel d'enseignants

Les données générées par ces items ont servi à réaliser les analyses quantitatives de l'étude. Toutefois, afin de réaliser les analyses qualitatives, un second instrument a dû être créé de toute pièce aux fins de la recherche. Cet instrument prend la forme d'un questionnaire adressé au panel d'enseignants de mathématiques et sert à recueillir leurs commentaires au sujet de deux items tirés du questionnaire contextuel des enseignants de la TEIMS de 2003. Ces deux items représentent des exemples d'items portant spécialement sur la pratique de pédagogies novatrices recommandées par les récents courants de réforme scolaire (MEFO, 1997; NCTM, 1999). Ces deux items ont été choisis puisqu'ils représentent des composantes importantes des nouvelles recommandations du curriculum de l'Ontario. Une étude de Masini et Taylor (2000) a pu établir un lien significatif positif précisément entre ces deux items et le niveau de rendement des élèves dans des écoles américaines. Ces deux items tentent de rejoindre des pratiques qui ciblent et augmentent la compréhension profonde des concepts mathématiques. Les items qui suivent ont été présentés au panel d'enseignants sur une fiche item incluse dans la trousse de sondage.

Item 1 :

Pendant les cours de mathématiques, vous arrive-t-il de demander aux élèves de réaliser les activités suivantes ?

- Expliquer le raisonnement qui sous-tend une idée.

CHOIX DE RÉPONSE : jamais ou presque jamais
 parfois
 la plupart du temps
 à chaque fois

Item 2 :

À quelle fréquence demandez-vous aux élèves de mathématiques de la classe TEIMS de faire ce qui suit?

- Écrire des équations et des fonctions qui représentent des relations.

CHOIX DE RÉPONSE : tous les cours
 la moitié des cours
 quelques cours
 jamais

Les questions qui s'y rapportent ont été élaborées dans le but de mesurer des concepts importants permettant de recueillir de l'information sur la qualité psychométrique de ces deux items. Les questions adressées aux enseignants du panel s'organisent en fonction des concepts psychométriques et des critères retenus aux fins de l'étude et sont présentées dans le tableau 5. Les questions intégrales se trouvent à l'annexe D du présent document.

Tableau 5

Questions adressées au panel d'experts

Concepts		
psychométriques	Critère évalué	Questions au panel
Validité de contenu	Formulation de la question	1) Que pensez-vous de la formulation de la question ?
		2) La question est-elle rédigée en respectant les règles de grammaire et de syntaxe ?
		3) Est-ce que la longueur de la question est acceptable ?
	Niveau de familiarité du vocabulaire utilisé	1) Est-ce que le vocabulaire vous est familier ?
		2) Est-ce que vous utilisez ce vocabulaire dans le cadre de votre travail ?
		3) Quel sens accordez-vous aux termes importants de la question ?
		4) Selon vous, est-ce que ces termes possèdent une définition généralement bien établie et uniforme auprès de vos collègues ?
	Clarté de la question	1) Est-ce que cette question est facilement comprise ?
		2) Est-ce que l'information demandée est claire pour vous ?
		3) Est-ce que vous croyez que la question pourrait être interprétée de différentes façons ?
Présence de plus d'un concept dans la question	1) Est-ce que la question cible plus d'un concept ? Si oui, selon vous lesquels ?	
	2) Est-ce que la présence de plus d'un concept agit sur la clarté de la question ?	

Tableau 5 (suite)

Validité de contenu (suite)	Pertinence pour le contexte de réforme de l'Ontario	<p>1) D'où proviennent vos connaissances en lien avec le sujet de la question (cours universitaires, atelier à l'école, documents du ministère, etc.) ?</p> <p>2) Est-ce que vous croyez que le sujet abordé par la question est important pour le contexte scolaire des francophones de l'Ontario ?</p> <p>3) Pensez-vous que le sujet de la question pourrait avoir un lien avec le rendement des élèves ?</p>
Biais de réponse	Présence d'un biais de réponse	<p>1) Est-ce que vous pouvez déceler un sens moral à cette question?</p> <p>2) Seriez-vous tenté de répondre d'une façon donnée si vous sachiez que votre directeur ou vos collègues avaient accès à vos réponses ?</p> <p>3) Selon vous, est-ce que votre école, votre conseil ou le ministère ont véhiculé des messages précis en lien avec le sujet abordé par la question?</p>
Validité de conséquence	Pertinence des choix de réponse	<p>1) Est-ce que vous trouvez que les choix de réponse sont adéquats?</p> <p>2) Est-ce que les choix de réponse vous permettent de traduire votre réalité de façon fidèle ?</p> <p>3) Est-ce que les choix de réponse vous permettent de répondre de façon suffisamment précise ?</p>

Procédure

Une demande a d'abord été présentée au comité de déontologie de l'Université d'Ottawa dans le but d'obtenir l'autorisation officielle pour procéder à l'étude. Entre temps, la sélection des items de l'étude fût effectuée parmi les questionnaires contextuels administrés aux enseignants et aux élèves lors des trois cycles de la TEIMS. Une fois l'approbation reçue du Bureau d'éthique en recherche de l'Université d'Ottawa, les fichiers de données pour les trois cycles d'enquête ont été récupérés directement à partir du site Internet de la TEIMS (<http://timss.bc.edu/#>). Des démarches ont ensuite été entreprises auprès des responsables de la TEIMS afin d'obtenir les codes d'identification pour les écoles de langue française de l'Ontario. Les fichiers de données et la majorité des analyses statistiques ont été réalisés à l'aide du logiciel SPSS, version 11.0 (2001). L'étape de préparation des fichiers comprenait notamment : la suppression de variables non pertinentes à l'étude, le jumelage des données des enseignants avec celles de leurs élèves respectifs et l'élimination de dossiers ayant trop de données manquantes.

Suite à l'obtention des permissions du Bureau d'éthique en recherche de l'Université d'Ottawa, les démarches de recrutement servant à recueillir les données qualitatives ont été entamées auprès de trois conseils scolaires francophones de l'Ontario. À chaque étape du processus, il était difficile de convaincre les participants. Toutes les directions d'écoles ont affirmé que les enseignants étaient très occupés et qu'ils n'osaient plus en demander davantage à ces personnes déjà surchargées – et, ce en dépit de leur intérêt pour le sujet de la recherche. Les données recueillies dans le cadre de ces efforts de recrutement ont ensuite été compilées et résumées.

Traitement des données

Le traitement des données de cette étude comporte des analyses de types quantitatif et qualitatif. Plusieurs des critères sélectionnés dans le cadre de la présente étude sont évalués à partir des données recueillies auprès du panel d'enseignants de mathématique. Leurs réponses aux questions présentées dans le tableau 5 ont d'abord été compilées et les opinions, partagées par plusieurs, ont ensuite été résumées et présentées dans la section des résultats. Les autres critères ont fait l'objet d'analyses variées. La présente section expose les analyses réalisées en fonction des concepts mesurés dans le but d'établir le degré de qualité psychométrique des items sélectionnés pour l'étude. Le tableau 6 illustre ces concepts, les critères et leurs méthodes d'évaluation. En plus des critères de qualité psychométrique, certaines analyses statistiques de bases ont été effectuées pour décrire d'abord les individus faisant partie des trois échantillons de données TEIMS (1995-1999-2003). Des analyses descriptives ont également été réalisées dans le but de décrire certaines des propriétés métriques associées aux résultats générés par les items des questionnaires, on retrouve notamment : les taux de réponse aux items, les mesures de tendance centrale, les mesures de dispersion, les coefficients d'asymétrie et de kurtose ainsi que des indicateurs de distribution normale. Les fruits de ces analyses préliminaires sont présentés au début de la section des résultats.

Tableau 6

Concepts, critères et méthodes d'analyse

Concepts	Critères évalués	Méthodes d'analyses
	Formulation de la question	Questions au panel
Validité apparente	Clarté de la question	Questions au panel
et	Présence de définition pour le vocabulaire utilisé	Examen des questionnaires contextuels
Validité de contenu	Niveau de familiarité du vocabulaire utilisé	Questions au panel
	Présence de plus d'un concept dans la question	Questions au panel
	Pertinence du sujet abordé par l'item pour le contexte éducatif	Questions au panel Tableau de spécifications
	Pertinence des choix de réponse	Questions au panel
	Présence et nombre de sous-questions	Examen des questionnaires contextuels
	Période de référence de l'item	Examen des questionnaires contextuels
	Sensibilité des items	Analyses de variance (ANOVA)

Tableau 6 (suite)

Concordance élève/enseignant	Degré de concordance entre les réponses aux items portant sur les nouvelles pratiques pédagogiques issus des questionnaires contextuels élève et enseignants de la TEIMS	Calculs de déviations moyennes entre les résultats des enseignants et de leurs élèves
Validité théorique	Établissement du construit en fonction des trois compétences pédagogiques	Cohérence des modèles exploratoires issus des analyses factorielles (AFE)
Biais de réponse	Évaluation de la présence d'un biais de réponse dû au phénomène de désirabilité sociale	Questions au panel
Fidélité	Fidélité des résultats et des indices composés	Analyses des covariances (cohérence interne)

En ce qui concerne la première catégorie du tableau 6, soit les concepts de *validité apparente et de contenu*, les critères suivants ont été étudiés par le panel d'enseignants : la formulation de la question, la clarté de la question, le niveau de familiarité du vocabulaire utilisé, la présence de plus d'un concept dans la question, la pertinence du sujet abordé par l'item pour le contexte éducatif et la pertinence des choix de réponse. Trois autres critères ont ensuite fait l'objet d'examen des questionnaires contextuels par le chercheur. D'abord, l'évaluation de la *présence de définition pour le vocabulaire utilisé* visait à déterminer s'ils contiennent ou non des définitions ou autre information pouvant aider les enseignants à

comprendre le sens des termes utilisés, surtout ceux qui représentent des pratiques novatrices. Le critère suivant servait à préciser la *présence et le nombre de sous-questions*. Si les questionnaires contextuels de la TEIMS en contiennent un grand nombre ceci peut avoir un impact sur le niveau de compréhension des items par les répondants. La *présence et le rang de la sous-question* seront rapportés pour chacun des items. Enfin, pour évaluer la *période de référence de l'item*, il s'agissait de déterminer, pour chacun des items, si la période de référence était clairement énoncée dans le questionnaire contextuel.

La sensibilité des items est évaluée par l'entremise d'analyses de variance (ANOVA). Ces analyses permettent d'identifier si des différences statistiquement significatives existent entre les fréquences moyennes observées pour les trois échantillons de l'étude (1995-1999-2003). Pour réaliser les analyses de variance, des items qui se répètent dans les questionnaires des trois cycles d'enquête ont d'abord été identifiés. Dans la pratique, l'ANOVA est une procédure statistique très robuste (Howell, 1998). Cependant, elle peut être compromise si les conditions d'application sont enfreintes de façon extrême. La première analyse de variance, dite *a priori* identifie les items pour lesquels il existe une différence significative entre les moyennes des trois cycles, sans toutefois préciser entre quels cycles. L'analyse ANOVA *post-hoc* est réalisée suite à cette première phase d'analyses et permet de repérer exactement où se trouvent les différences significatives. La méthode du « test de la plus petite différence significative de Fisher », mieux connue sous l'acronyme LSD (Least Square Difference), est utilisée pour produire ces résultats. Ces analyses sont suivies d'une analyse de variance non-paramétrique (Kruskal-Wallis) dans le but de confirmer les résultats obtenus par les méthodes paramétriques.

En ce qui a trait au deuxième concept du tableau 6, *la concordance entre élève et enseignant*, les résultats de chaque enseignant ont été comparés aux résultats de leurs élèves respectifs. Lors d'analyses secondaires menées sur des données d'évaluation à grande échelle, il n'est pas possible de faire un retour en arrière pour inclure le jugement d'un observateur ou encore pour une deuxième prise de données dans le but d'investiguer un potentiel biais de réponse. Puisque les items varient quelquefois entre les cycles d'enquête, le mariage des items enseignant-élève a été effectué de façon indépendante d'un cycle à l'autre. Cinq items furent jumelés entre les enseignants et leurs élèves pour les échantillons de 1995 et de 1999 et de dix items pour 2003. Il fallait d'abord créer des scores standardisés pour l'ensemble des données. En effet, comme certaines échelles des variables varient entre elles (trois à cinq intervalles), les comparaisons peuvent être faites plus facilement en calculant une cote Z pour chaque score avant d'effectuer les comparaisons. Ensuite, le calcul des moyennes fût effectué pour tous les items des élèves, et ce, pour chacune des classes. La deuxième étape consistait à calculer la différence entre le résultat de chaque élève et la moyenne de sa classe (en valeur absolue), et ce, pour chaque item. Pour obtenir la déviation moyenne des élèves, il était nécessaire d'additionner toutes ces différences et de diviser cette somme par le nombre d'élèves dans la classe. Ce calcul est fait dans le but de vérifier la qualité des résultats du questionnaire des élèves. Si les items du questionnaire génèrent des réponses valides, ceci devrait être reflété dans le degré de fidélité des réponses entre les élèves d'une même classe. L'étape suivante demandait de nouveau un calcul de différence en valeur absolue, mais cette fois, entre les scores des élèves et celui de leur enseignant. La moyenne des valeurs obtenues est calculée afin de produire un score unique pour chaque item sélectionné qui indiquera la déviation moyenne des réponses entre l'enseignant et ses élèves.

Afin d'analyser la *validité théorique*, le troisième concept du tableau 6, des analyses factorielles exploratoires (AFE) sont réalisées avec comme objectif de déterminer si les items de l'étude se regroupent conformément aux compétences mathématiques recommandées par le MEFO. Si les résultats des analyses factorielles démontraient une structure comparable au regroupement effectué préalablement en fonction du contenu théorique, cela pourrait indiquer que les items représentent en effet des mesures d'un même concept théorique sous-jacent. L'utilisation du logiciel Mplus (Muthen & Muthen, 1998) offre un estimateur WLSMV (weighted least-squares with mean and variance adjustment) suggéré pour une utilisation avec des données catégorielles. Une rotation oblique (promax) a été retenue pour l'ensemble des analyses permettant une corrélation entre les différents facteurs.

Avant de mener les analyses factorielles, quelques éléments ont dû être considérés. Les analyses factorielles ont été réalisées en deux phases. On doit d'abord rappeler que lors de la création des échantillons finaux, les participants (enseignants) ayant 30% ou moins de données manquantes ont été retenus. La procédure factorielle requiert un échantillon complet sans donnée manquante. Lors de la première phase, les données manquantes ont reçu un traitement de suppression « listwise » où le logiciel élimine un enregistrement complet lorsqu'il contient au moins une donnée manquante. Pour l'échantillon de 1995, les 123 participants de l'échantillon ont été réduits à 94 suite à ce traitement, une perte de 23,6% de l'échantillon. Pour 1999, le nombre de 90 participants a chuté à 75 (-16,6%), et pour 2003 des 75 enregistrements, il n'en restait que 28 suite à la procédure (-62,6%). Pour l'échantillon de 2003, six variables ayant un taux de non-réponse plus élevé furent éliminées et cela a permis de conclure la première ronde des analyses avec un échantillon de 59

enseignants. Lors des analyses, les variables ont uniquement été retenues pour un facteur lorsque le coefficient associé affichait une valeur supérieure à 0,4. Afin d'obtenir des structures simples, une autre recommandation fût adoptée, soit la retenue des variables affichant un seul coefficient associé supérieur à 0,4 pour une solution donnée. De plus, deux indices d'ajustement ont servi à obtenir une mesure de l'harmonisation du modèle théorique (exploratoire) avec les données empiriques de l'étude : RMSEA (*Root Mean Square Error of Approximation*) et RMSR (*Root Mean Square Residual*). Selon Hu et Bentler (1999), ces indicateurs devraient afficher des valeurs respectivement inférieures à 0,08 et 0,06.

Afin d'analyser la possibilité d'un *biais de réponse*, au cinquième concept du tableau 6, les enseignants du panel ont fourni leurs opinions quant à la présence du phénomène de désirabilité sociale chez les répondants. En réponse aux diverses pressions sociales et professionnelles, les enseignants peuvent rapporter une utilisation plus fréquente des nouvelles pédagogies. Un biais systématique de réponse risque de teinter les résultats générés par les instruments de mesure. Les propos des membres du panel sont résumés et présentés dans la section des résultats.

Enfin, les analyses liées à la fidélité des variables composées portent sur le dernier concept du tableau 6. Dans le cadre de cette étude, la fidélité des items de la TEIMS est évaluée selon la méthode des covariances. À l'aide du coefficient alpha de Cronbach, une estimation de la proportion de variance due au score vrai est obtenue, ce qui est communément appelé la cohérence interne ou de fidélité. Les items sont d'abord regroupés en fonction du concept qu'ils sont soupçonnés de mesurer, soit : le raisonnement mathématique, la résolution de problème ou la communication mathématique ou globalement, les pratiques pédagogiques nouvelles. Les analyses de cohérence interne ont

été lancées à partir de ces premiers regroupements basés sur une analyse de contenu théorique des items avec le logiciel SPSS, version 11.0 (2001).

En résumé, les participants de l'étude sont des enseignants de mathématiques de 7^e et de 8^e année qui oeuvrent dans des écoles de langue française de l'Ontario ainsi que leurs élèves respectifs. Ils ont tous répondu à un questionnaire contextuel dont la présente étude examine certains items portant sur les pratiques pédagogiques des réformes. Des commentaires additionnels ont été recueillis à l'aide d'un sondage administré à un panel d'enseignants de mathématiques francophones de l'Est ontarien. Le traitement des données, une série d'analyses qualitatives et quantitatives, fût élaboré en fonction des concepts de la qualité psychométrique et se voit appliqué aux données obtenues à partir des questionnaires contextuels de la TEIMS et du sondage.

Résultats

Cette section décrit tous les résultats des analyses réalisées sur les échantillons de données présentés dans la section précédente. Les résultats exposés dans cette section s'organisent selon sept catégories : 1) caractéristiques démographiques des échantillons, 2) propriétés métriques des variables de l'étude, 3) validité apparente et de contenu, 4) concordance entre élève et enseignant, 5) validité théorique, 6) biais de réponse et 7) fidélité. La présente section contient les résultats des analyses qualitatives et quantitatives visant à évaluer la qualité psychométrique des items à l'étude.

Caractéristiques démographiques des échantillons

Suite à la sélection finale des trois échantillons de données de la TEIMS (1995-1999-2003), des analyses descriptives ont été réalisées dans le but de dresser un portrait global des principales caractéristiques démographiques des échantillons. Cet exercice sert à déterminer l'adéquation de l'échantillon à l'ensemble de la population visée par l'étude. D'abord, le tableau 7 démontre qu'un plus grand nombre de femmes composent les trois échantillons. Ce résultat reflète les propos de l'Ordre des enseignantes et des enseignants de l'Ontario qui rapporte, que les enseignants masculins représentent uniquement 31 % de tous les enseignants membres du regroupement et que la situation est davantage asymétrique dans les écoles françaises de l'Ontario, où les hommes ne représentent que 8 % des enseignants juniors au primaire (Giguère, 1999).

Tableau 7

Répartition selon le sexe des enseignants dans les trois échantillons

Sexe	1995		1999		2003	
	Fréquence	%	Fréquence	%	Fréquence	%
Femme	75	60,98	52	57,78	47	62,67
Homme	47	38,21	38	42,22	28	37,33
Données non disponibles	1	0,80	0	--	0	--
Total	123	100	90	100	75	100

Une analyse descriptive ciblant l'âge des participants permet de voir que leur répartition varie grandement d'un échantillon à l'autre (voir tableau 8). Dans le premier échantillon, soit celui de 1995, presque la moitié des participants sont âgés entre 40 et 49 ans, tandis qu'ils ne représentent qu'environ 10% de l'échantillon de 2003. On constate également une augmentation marquée du pourcentage d'enseignants dans la catégorie des 30 à 39 ans, en effet ce pourcentage double entre les échantillons de 1995 et 2003. De plus, il importe de souligner que le nombre d'enseignants de 25 à 29 ans dans l'échantillon de 2003 est presque trois fois supérieur au même groupe d'âge pour les échantillons de 1995 et 1999.

En ce qui concerne l'expérience des enseignants en exercice, les résultats rapportés dans le tableau 9 démontrent qu'en moyenne, les enseignants de l'échantillon de 2003 possèdent moins d'expérience que les autres participants de l'étude; le nombre moyen d'années d'expérience étant presque deux fois plus petit en 2003 qu'en 1995 et 1999.

Tableau 8

Répartition selon l'âge des enseignants dans les trois échantillons

Catégories d'âge	1995	1999	2003
	%	%	%
Moins de 25 ans	4,07	2,22	2,67
25 à 29 ans	9,76	7,78	26,67
30 à 39 ans	21,95	28,89	46,67
40 à 49 ans	47,15	30,00	10,67
50 à 59 ans	14,63	30,00	12,00
60 et plus ans	0,81	--	0,00
Données non disponibles	1,63	1,11	1,33
Moyennes d'âge	41,25	42,73	35,64
Écart type	8,98	8,19	8,87

Tableau 9

Années d'expérience des enseignants des trois échantillons

Années d'expérience	1995	1999	2003
Moyenne	17,15	17,06	9,73
Minimum	1	1	1
Maximum	34	35	31
Médiane	19	17	8

En ce qui concerne le niveau d'éducation des participants, les résultats sont résumés dans le tableau 10. D'abord, le pourcentage des enseignants n'ayant pas obtenu un diplôme de premier cycle diminue à chaque échantillon pour n'atteindre que 2,67% dans celui de 2003. De plus, une augmentation est notée en 2003 au niveau de la proportion des enseignants qui détiennent au minimum un B.A. avec ou sans formation à l'enseignement.

Tableau 10

Répartition selon le niveau d'éducation des enseignants dans les trois échantillons

Niveaux d'éducation	1995	1999	2003
	%	%	%
Études secondaires et 1 ou 2 ans de formation à l'enseignement			
Études secondaires et 3 ou 4 ans de formation à l'enseignement	13,12	6,67	2,67
B.A. ou équivalent sans formation à l'enseignement			
B. A. ou l'équivalent avec formation à l'enseignement	78,69	77,78	90,67
M.A. P.H.d. sans formation à l'enseignement			
M.A. P.H.d. avec formation à l'enseignement	8,20	15,56	6,67

Propriétés métriques des variables de l'étude

Le potentiel des items en tant que « générateurs d'information » représente une considération importante dans l'examen de leurs propriétés métriques. L'examen présenté porte notamment sur le type de variable générée (dichotomique, intervalle) et un ensemble de caractéristiques de distribution telles que les mesures de tendance centrale, les mesures de dispersion, les coefficients d'asymétrie et de kurtose, et la probabilité de normalité.

D'abord, le type de variable générée est examiné puisqu'il dicte souvent le choix des analyses statistiques réalisées et que les résultats des analyses statistiques se voient influencés par le type de variable. Les données de nature catégorique restreignent l'ampleur des effets décelés par l'application d'analyses statistiques basées sur les corrélations entre variables. Il est très difficile de connaître l'ordre de grandeur de la diminution de l'effet pour ce type de données. Le type de variable générée est évalué en fonction de la conception des items du questionnaire contextuel de la TEIMS. On constate notamment que la grande majorité des items retenus dans le cadre de l'étude produisent des variables catégorielles ayant entre trois et cinq catégories de réponse. Le tableau 11 présente les proportions de variables générées selon leur nombre de catégories de réponses pour les items des trois cycles d'enquête. Les items à quatre catégories de réponse représentent la plupart des items de l'étude.

Tableau 11

Pourcentage du type de variables générées par les questionnaires de la TEIMS

1995-1999-2003

	1995	1999	2003
Type de variable générée	%	%	%
2 catégories	0	0	4
3 catégories	26	30	19
4 catégories	70	70	73
5 catégories	4	0	0
Variables continues	0	0	4

Ensuite, tous les items retenus à partir des questionnaires des enseignants de la TEIMS ont fait l'objet d'une analyse pour déterminer leur taux de réponse. Ce taux est calculé en divisant le nombre de participants ayant fourni une réponse valide dans les données rendues publiques par l'organisation de l'IEA par le nombre total de participants. Cette analyse fournit des pistes servant à l'identification d'items généralement problématiques pour les répondants. Il est souhaitable d'obtenir un taux de réponse élevé et uniforme pour chacun des items d'un test ou questionnaire. Tel que mentionné déjà, lors de la création des fichiers d'analyse, les échantillons ont été restreints afin d'éliminer les cas qui comportaient un trop grand nombre de données manquantes. Si les analyses tenaient compte uniquement des taux de réponse pour les cas retenus dans l'échantillon final, une partie importante de l'information serait perdue. C'est pourquoi, les analyses ont été réalisées séparément sur les deux échantillons (avant suppression des données manquantes et après), et ce, pour chaque cycle d'enquête (1995, 1999, 2003). Donc, les analyses ont d'abord été réalisées sur les échantillons originaux, c'est-à-dire ceux contenant tous les dossiers TEIMS disponibles pour les enseignants de mathématiques des écoles françaises de l'Ontario. Les mêmes analyses ont ensuite été menées sur les échantillons finaux, i.e. les échantillons réduits suite à la suppression de certains enregistrements ayant trop de données manquantes (voir section du traitement des données). La figure 2, réalisée à partir des données originales, présente les taux de réponse minimum et maximum pour l'ensemble des items retenus pour l'étude. Les taux maximum sont plus élevés pour les cycles de 1995 et de 2003. Le taux minimum de 2003 tombe sous le seuil critique du 50%. En réalité, l'échantillon de 2003 compte quatre items dont le taux de réponse tombe sous la barre des

50%. Tous ces items commencent par la même formulation, soit : « À quelle fréquence utilisez-vous l'ordinateur pour les activités ci-dessous, dans vos cours de mathématiques de la classe TEIMS? » l'activité est ensuite précisée. Les taux de réponse pour chacun des items sont inclus dans un tableau à l'annexe G.

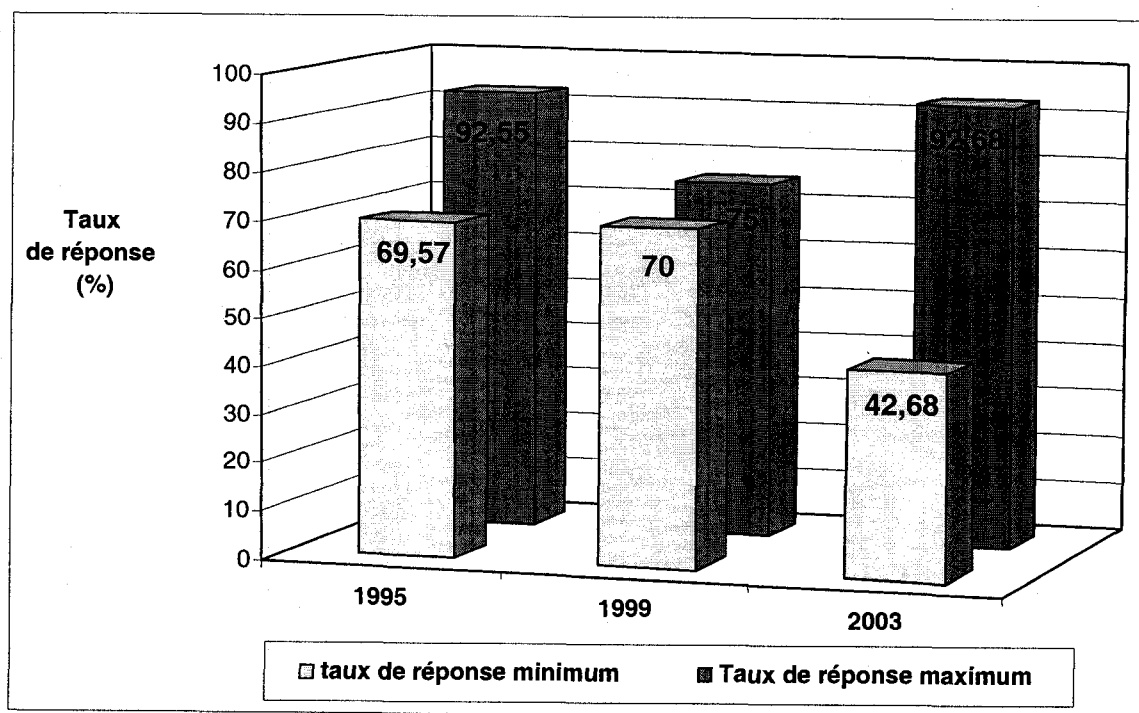


Figure 2. Taux de réponse minimum et maximum aux items du questionnaire enseignant – pour les trois cycles - données originales

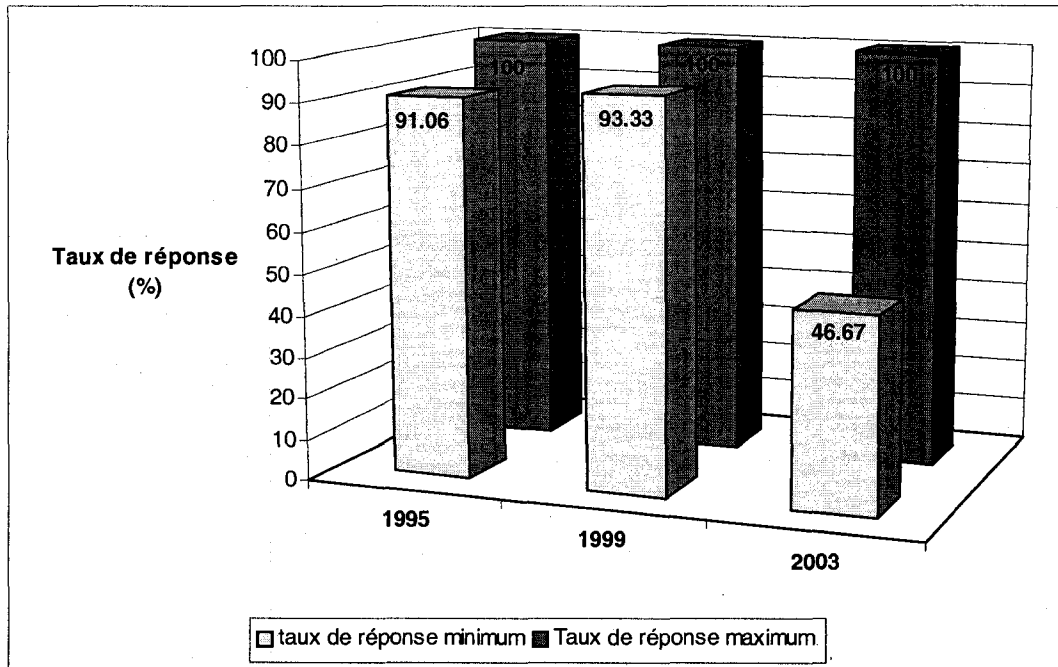


Figure 3. Taux de réponse minimum et maximum aux items du questionnaire enseignant – pour les trois cycles - données finales

La figure 3 présente le même portrait, mais cette fois pour les échantillons finaux, c'est-à-dire les fichiers réduits obtenus suite à l'élimination des certaines données manquantes. Clairement, elle démontre à quel point l'élimination de certains dossiers a modifié les échantillons. Cette étape fût nécessaire afin de pouvoir mener des analyses statistiques exigeant un plus grand nombre de données valides. L'élimination de dossiers n'a pas trop modifié le taux de réponse minimum pour les résultats de 2003, où les même quatre items discutés précédemment sont toujours en cause.

En éliminant les quatre items problématiques de l'échantillon de 2003, alors les moyennes des trois échantillons deviennent vraiment très semblables, puisque le taux moyen pour 2003 serait à ce moment équivalent à 98,00%. On remarque dans les tableaux de l'annexe G que les items issus d'un même énoncé de question possèdent des taux de réponse quasi-identiques.

L'examen des propriétés métriques comprend également les mesures de tendance centrale, soit la moyenne, la médiane et le mode. Pour la grande majorité des items de l'étude les données sont soit catégorielles ou ordinales. À la limite, il serait possible d'argumenter que les données sont ordinales à intervalles réguliers, ce qui leur donnerait une qualité semblable à celle des données continues. Il est important de comprendre que les mesures de tendance centrale rapportées dans le cadre de cette étude, servent uniquement à fournir leur part d'information quant à la distribution (asymétrie surtout) des variables. Afin de repérer rapidement les items pouvant s'avérer problématiques, les moyennes situées dans la zone centrale de l'échelle de réponse, sont identifiées comme étant acceptables. Donc, les items affichant une moyenne située entre 25% et 75 % de l'espace compris sur l'échelle des options de réponse sont acceptables. La figure 4 illustre cette règle de décision par la zone centrale ombragée. Les items pour lesquels la moyenne calculée figure à l'extérieur de la zone centrale indiquée par la figure 4 apparaissent en surbrillance dans les tableaux 15, 16 et 17.

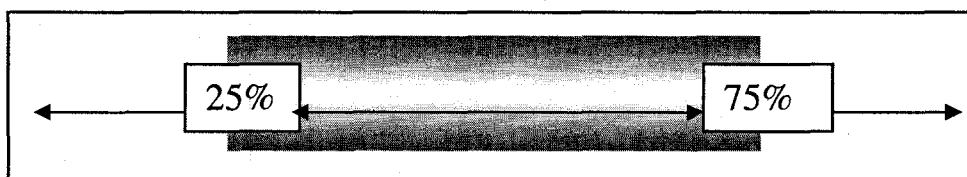


Figure 4, Illustration des bornes servant à identifier les items ayant des moyennes problématiques

Tableau 12

Moyenne, médiane et mode des items enseignants de l'échantillon de 1995

Items	Nombre de données		Échelle	Moyenne	Médiane	Mode
	Valides	Manquantes				
Item 1	123	0	3	2,28	2	2
Item 2	122	1	4	3,75	4	4
Item 3	116	7	4	2,33	2	2
Item 4	123	0	4	2,87	3	3
Item 5	123	0	5	4,52	5	5
Item 6	122	1	4	1,96	2	2
Item 7	123	0	4	3,16	3	4
Item 8	122	1	4	2,50	3	3
Item 9	123	0	4	1,98	2	2
Item 10	123	0	4	1,79	2	2
Item 11	123	0	4	1,46	1	1
Item 12	123	0	3	2,93	3	3
Item 13	119	4	4	3,23	3	3
Item 14	122	1	4	2,80	3	4
Item 15	120	3	4	2,09	2	2
Item 16	112	11	4	2,38	2	3
Item 17	118	5	4	2,43	2	2
Item 18	123	0	3	2,94	3	3
Item 19	123	0	3	2,96	3	3
Item 20	123	0	3	2,68	3	3
Item 21	119	4	4	2,97	3	3
Item 22	116	7	4	2,30	2	3
Item 23	123	0	3	2,89	3	3

Note : La colonne « échelle » précise le nombre de catégories de réponse.

Tableau 13

Moyenne, médiane et mode des items enseignants de l'échantillon de 1999

Items	Nombre de données		Échelle	Moyenne	Médiane	Mode
	Valides	Manquantes				
Item 1	90	0	3	2,54	3	3
Item 2	89	1	4	3,54	4	4
Item 3	85	5	4	2,31	2	2
Item 4	89	1	4	2,88	3	3
Item 5	90	0	3	2,26	2	2
Item 6	90	0	4	2,16	2	2
Item 7	87	3	4	3,53	4	4
Item 8	90	0	4	2,41	2	2
Item 9	89	1	4	2,07	2	2
Item 10	90	0	4	1,92	2	2
Item 11	90	0	4	1,34	1	1
Item 12	90	0	3	2,74	3	3
Item 13	85	5	4	3,00	3	3
Item 14	86	4	4	3,15	3	4
Item 15	90	0	4	2,02	2	2
Item 16	89	1	4	2,24	2	3
Item 17	90	0	4	2,48	2	2
Item 18	90	0	3	2,87	3	3
Item 19	90	0	3	2,88	3	3
Item 20	90	0	3	2,66	3	3
Item 21	84	6	4	2,82	3	3
Item 22	89	1	4	2,46	3	3
Item 23	90	0	3	2,79	3	3

Note : La colonne « échelle » précise le nombre de catégories de réponse.

Tableau 14

Moyenne, médiane et mode des items enseignants de l'échantillon de 2003

Items	Nombre de données		Échelle	Moyenne	Médiane	Mode
	Valides	Manquantes				
Item 1	77	0	4	1,62	2	2
Item 2	77	0	4	3,50	4	4
Item 3	74	3	4	2,68	3	3
Item 4	76	1	4	3,45	4	4
Item 5	76	1	3	2,25	2	2
Item 6	76	1	4	2,11	2	2
Item 7	71	6	4	3,30	3	4
Item 8	75	2	4	2,20	2	2
Item 9	76	1	4	2,49	2	2
Item 10	76	1	4	2,80	3	3
Item 11	37	40	4	1,81	2	2
Item 12	76	1	4	2,83	3	3
Item 13	77	0	4	3,24	3	3
Item 14	70	7	4	2,87	3	4
Item 15	76	1	5	1,96	2	1
Item 16	75	2	3	2,11	2	2
Item 17	75	2	3	2,74	3	3
Item 24	77	0	4	3,01	3	3
Item 25	77	0	4	3,55	4	4
Item 26	77	0	99	18,18	18,5	10
Item 27	76	1	4	2,83	3	2
Item 28	77	0	2	1,47	1	1
Item 29	36	41	4	1,80	2	2
Item 30	37	40	4	1,78	2	2
Item 31	37	40	4	1,86	2	2
Item 32	74	3	3	2,12	2	2

Note : *La colonne « échelle » précise le nombre de catégories de réponse.

En somme, un grand nombre d'items ont été identifiés car leur moyenne se situe à l'extérieur de la zone acceptable. Il est nécessaire de préciser pourquoi cette information importe dans le présent contexte et comment celle-ci a été interprétée. Les items de questionnaires sont souvent utilisés par les chercheurs et les analystes dans le cadre d'analyses à base corrélacionnelle. Cependant, sans variance dans les résultats des questionnaires, il est impossible d'établir des corrélacions statistiques avec d'autres variables clés, telles que le rendement des élèves. En effet, dans un climat post-réforme, les chercheurs peuvent s'attendre à ce que certaines pratiques pédagogiques récentes soient très répandues chez les enseignants, ce qui, avec les items dont on dispose, produirait une moyenne élevée, une faible dispersion et une asymétrie négative. Ces réalités font en sorte que même si les items de questionnaires sont généralement conçus autour de concepts valables aux yeux des chercheurs, la quantité d'information qu'ils produisent s'avère décevante et très limitée. Souvent, les échelles de réponse des items restreignent sérieusement leur variance, ce qui occasionne un manque de sensibilité envers les changements réels qui s'opèrent au niveau des pratiques et des croyances pédagogiques des enseignants.

Au niveau des items de 1995, neuf items sur un total de 23 affichent des moyennes à l'extérieur des bornes. Une constatation importante s'impose, non seulement les items de 1999 affichent-ils un nombre d'items problématiques égal à ceux de 1995, mais tous ces items sont également des jumeaux de forme et de contenu. Afin de comparer les items jumeaux entre les cycles d'enquête, on peut se référer à l'annexe C qui contient un tableau de tous les items correspondants en forme et contenu. Pour 2003, dix items ont été identifiés comme étant problématiques à partir de la banque des 26 items. De ces dix items, trois items seulement ont également été identifiés dans les échantillons de 1995 et de 1999. Le premier

de ces items problématiques dans les trois échantillons est l'Item 2, qui traite du degré d'accord des enseignants avec l'utilisation de multiples modes de représentation pour l'enseignement des mathématiques. La question et les échelles de réponse varient très légèrement entre les cycles d'enquête pour cet item. L'item 5 s'avère être le deuxième item problématique dans les trois échantillons. Il porte sur l'utilisation de la calculatrice dans les cours de mathématiques. Les questions et les échelles de réponse varient un peu d'un cycle à l'autre même si le sujet ciblé demeure exactement le même. Le troisième et dernier item dont la moyenne est problématique dans tous les échantillons est l'item 7. Celui-ci rapporte la fréquence d'utilisation de la calculatrice chez les élèves pour résoudre des problèmes complexes dans les cours de mathématiques. Les questions demeurent presque identiques entre les cycles mais les échelles de réponse ont été modifiées légèrement entre les administrations.

De plus, à partir des items identifiés dans les tableaux 12, 13 et 14, une grande proportion de ceux-ci correspondent au même énoncé de question. En effet, dans les échantillons de 1995 et de 1999, 12 des 18 items problématiques représentent tous des sous-questions de l'énoncé principal qui suit : « Pour qu'ils réussissent en mathématiques à l'école, dans quelle mesure pensez-vous qu'il est important pour les élèves... ». Les participants peuvent choisir parmi trois choix de réponse, soit : pas important, plus ou moins important, très important. Tous ces items affichent des moyennes à l'extérieure de la zone centrale.

L'examen des propriétés métriques porte également sur les mesures de dispersion, soit l'étendue, l'écart type et la variance. Encore une fois, afin de retirer un maximum d'information par l'administration d'un item et pour que celui-ci puisse servir à décrire des individus distincts au sein d'une même population, il est préférable que l'étendue et l'écart type ne soit pas trop restreints. L'étendue qui représente la différence entre le score le moins élevé et le score le plus élevé, fait partie des statistiques utilisées puisqu'il s'agit d'une façon rapide de repérer les items pour lesquels certains choix de réponse auraient été ignorés par tous les répondants. L'examen de l'écart type et de la variance des variables fournit des précisions quant à la dispersion des données. L'information statistique est rapportée sous forme de graphiques dans le but de simplifier la compréhension d'un grand nombre de résultats. Cependant, il est recommandé de se référer à l'annexe J pour consulter les tableaux contenant toutes les données. Dans l'annexe J, les items pour lesquels plus d'un choix de réponse ont une fréquence de zéro sont identifiés en surbrillance. Les figures 5, 6 et 7 illustrent l'étendue et l'écart type de chaque item. Le trait mince noir indique l'étendue et la bande grise plus large représente l'écart type centré sur la moyenne. Un item optimal afficherait un étendu couvrant tous ses choix de réponse et un écart type centré et suffisamment réparti sur l'échelle de réponse. L'item 17 de la figure 5 est un bon exemple d'un item optimal.

En regardant les données de 1995, clairement les items 2, 11, 12, 18 et 19 se distinguent par leur étendue et/ou leur écart type restreints. Ces items ne peuvent fournir autant d'information que d'autres items du même genre puisque certains de leurs choix de réponse ne sont pas utilisés et parfois les réponses sont concentrées dans une même région de l'échelle. Pour 1999, il s'agit des items 1, 2, 4, 11, 13, 16, 18, 19 et 22. Pour l'échantillon de 2003, les items 1, 2, 4, 6, 7, 11, 17, 29, 30 et 31 s'avèrent problématiques. L'item 28 pour les données de 2003 est une variable dichotomique et donc même si son étendue paraît restreint sur l'illustration graphique, elle doit être considérée autrement.

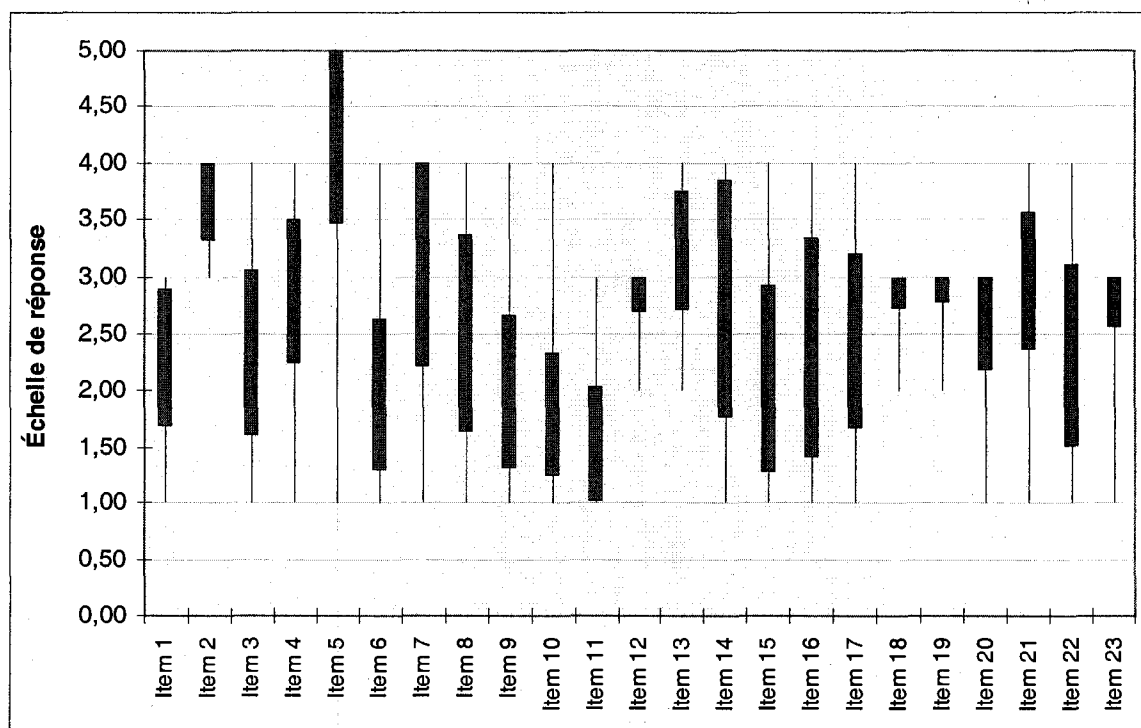


Figure 5. Étendues et écarts types des items de l'échantillon 1995

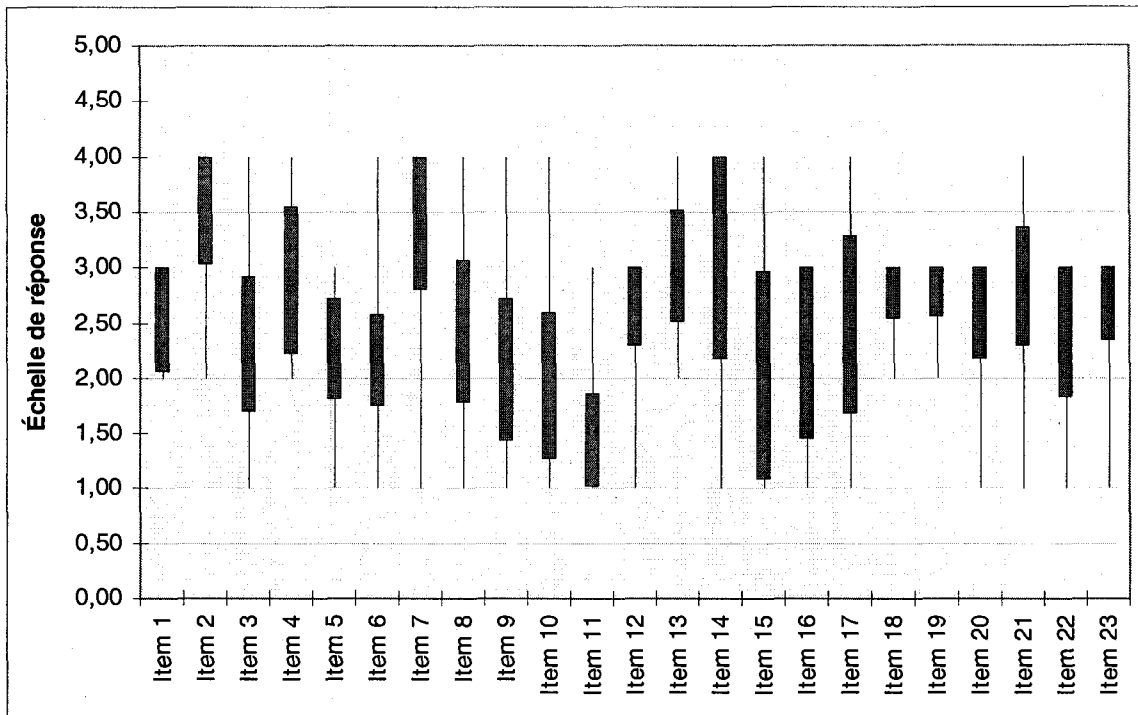
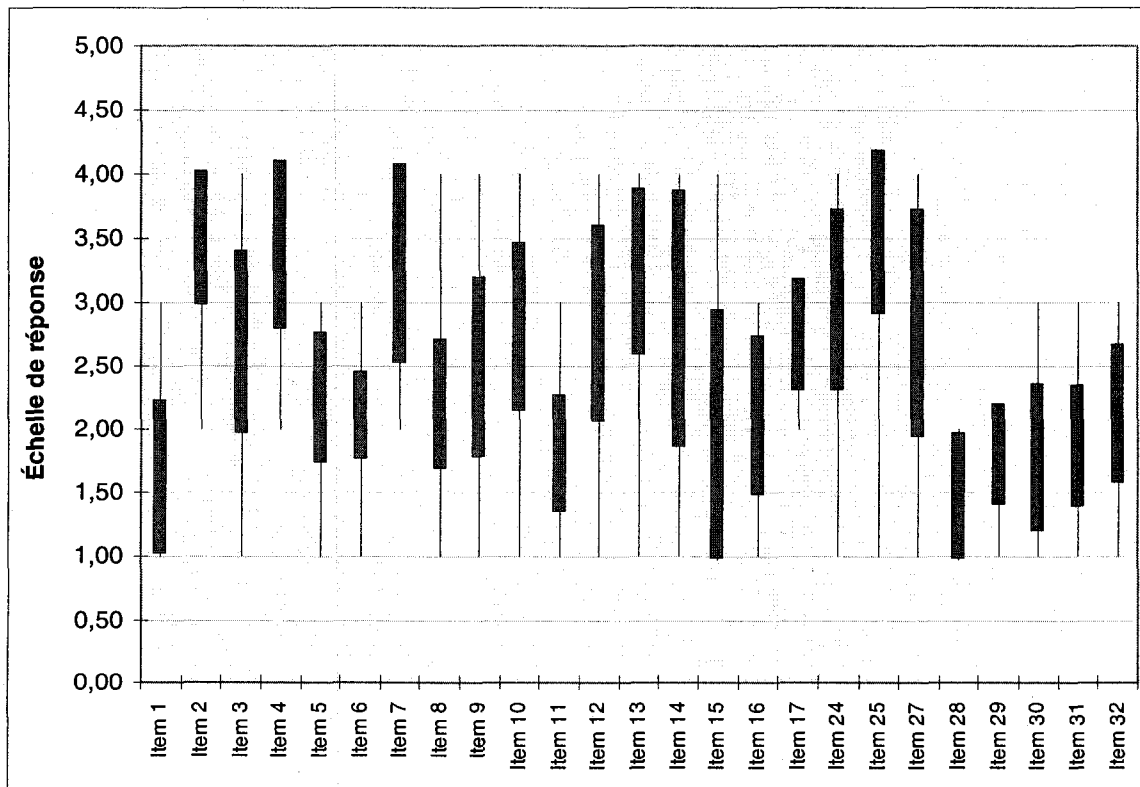


Figure 6. Étendues et écarts types des items de l'échantillon 1999



Note : l'item 26 est une variable continue et affiche une moyenne de 18,18 (écart type = 9,88) et ne pouvait figurer dans le même graphique puisque la différence était trop grande pour être représentée sur une échelle commune.

Figure 7. Étendues et écarts types des items de l'échantillon 2003

Tous les items de l'étude qui portent sur l'utilisation d'ordinateurs pour l'enseignement ou l'apprentissage des mathématiques ont des fréquences très faibles et souvent des choix de réponse ignorés dans l'extrémité supérieure de l'échelle de réponse (items 11 pour 1995, 1999 et 2003, items 29, 30, 31 pour 2003) . Plusieurs items au sujet de diverses croyances pédagogiques sont également problématiques au niveau des mesures de dispersion. Il s'agit principalement d'items formulés comme suit : « Pour qu'ils réussissent en mathématiques à l'école, dans quelle mesure pensez-vous qu'il est important pour les élèves de ... » ou « Dans quelle mesure êtes-vous d'accord avec chacun des énoncés suivants ? ». Pour 1995, quatre items problématiques sur cinq correspondent à ces formulations, et pour les données de 1999 ces items représentent cinq des neuf items problématiques.

Une autre composante importante de la description des propriétés métriques repose sur les coefficients d'asymétrie et de kurtose (ou voussure). La kurtose « renvoie fondamentalement à la concentration relative de scores au centre, dans les extrémités supérieure et inférieure (les queues), et au niveau des épaules (entre le centre et les queues) d'une distribution » (Howell, 1998, p.32). Tandis que le coefficient d'asymétrie précise à quel degré la distribution d'une variable est asymétrique, i.e. inégale par rapport à son centre. Ces coefficients sont produits pour l'ensemble des variables de l'étude et sont accompagnés de leur erreur type qui dépend principalement de la taille de l'échantillon. Pour une meilleure interprétation des résultats, l'erreur type de chacun des coefficients est utilisée pour générer deux indicateurs supplémentaires, soit la cote z du coefficient et un intervalle de confiance de 95%. Les tableaux de données contenant les coefficients et les indicateurs pour chacun des items figurent en annexe H. Ces résultats sont résumés dans le tableau 15 qui suit, où les X démontrent soit une kurtose ou une asymétrie problématique selon les

intervalles de confiance. La zone ombragée est utilisée pour montrer les items disponibles pour chacun des cycles. On peut constater que certains items ont à la fois une asymétrie et une kurtose inadéquate mais le lien entre les deux n'est pas direct. En tout, pour 1995, 11 items possèdent une asymétrie problématique, neuf items démontrent une kurtose qui soit différente de la distribution normale, et de ceux-là, sept items s'approprient ces deux caractéristiques à la fois. En ce qui concerne les données de 1999, 16 items affichent des asymétries problématiques, dix items des kurtoses inadéquates, dont sept items possédant les deux lacunes. Neuf items de 2003 représentent des asymétries importantes et dix items des kurtoses inadéquates, avec un total de six items ayant ces deux problématiques.

Tableau 15

Items de la TEIMS présentant une asymétrie ou une kurtose problématiques

N° item	1995		1999		2003	
	Asymétrie	Kurtose	Asymétrie	Kurtose	Asymétrie	Kurtose
Item 1				X		
Item 2	X			X		X
Item 3						
Item 4					X	
Item 5	X	X	X			
Item 6	X	X	X	X	X	X
Item 7	X		X		X	X
Item 8			X		X	X
Item 9			X	X		
Item 10		X	X	X		
Item 11	X		X			
Item 12	X	X	X			
Item 13			X	X	X	X
Item 14		X				X
Item 15			X		X	
Item 16			X			
Item 17				X	X	
Item 18	X	X				
Item 19	X	X	X	X		
Item 20	X		X	X		
Item 21	X	X	X			
Item 22			X	X		
Item 23	X	X	X			
Item 24						
Item 25					X	X
Item 26					X	X
Item 27						
Item 28						X
Item 29					X	
Item 30						
Item 31						
Item 32						

Pour conclure la section sur les propriétés métriques, les résultats des tests de Kolmogorov-Smirnov et de Shapiro-Wilk sont présentés. Ces tests sont conçus pour évaluer la probabilité que la distribution des items soit normale dans la population. Ces deux tests ont été réalisés sur les variables des trois cycles d'enquête et les résultats sont uniformes. Les deux tests ont produit les mêmes jugements pour l'ensemble des variables, i.e. qu'aucune variable ne serait distribuée normalement dans la population. Même si l'utilisation de ces tests est recommandée pour des variables catégorielles, il est difficile de savoir si le nombre restreint de choix de réponse a pu influencer les résultats. Les résultats complets sont présentés en annexe I.

Validité d'apparence et de contenu

Tel que présenté dans la section de la méthodologie, les analyses liées à la démonstration des concepts de validité d'apparence et de contenu regroupent diverses composantes. La prochaine section comporte les résultats des analyses réalisées pour chaque critère évalué en fonction de ces types de validité. La section comprend à la fois des résultats découlant des échantillons de la TEIMS et d'autres qui proviennent de l'administration des questions au panel d'enseignants.

Pour évaluer la *formulation de la question*, les enseignants du panel ont répondu à trois questions traitant de la formulation des questions exemplaires issues des questionnaires de la TEIMS. En ce qui a trait au premier item TEIMS, les opinions du panel sont partagées. Cet item demande à quelle fréquence les élèves doivent expliquer leur raisonnement. Certains enseignants du panel perçoivent la formulation de la question comme étant très bien, tandis que d'autres y voient des ambiguïtés et trouvent la formulation trop vague. Un paneliste précise qu'il a du recourir au dictionnaire pour être certain de bien comprendre la

question. Les opinions divergent également en ce qui concerne le respect des règles de grammaire et de syntaxe. Puisque l'item de la TEIMS est présenté au panel sur une fiche, et non dans son état original, i.e. comme un item inséré parmi d'autres dans un questionnaire, certains participants s'interrogent sur le recours du style télégraphique pour exprimer une seule activité. Ils précisent que l'énoncé ne devrait pas être utilisé pour introduire une seule amorce de question, ce qui en réalité, ne se produit pas dans le questionnaire de la TEIMS. Il y a consensus au niveau de la longueur de la question, c'est-à-dire que tous les commentaires reçus à cet effet indiquent une longueur acceptable. Le participant affichant le plus grand nombre d'années d'expérience, soit 27 ans, inclut un commentaire à l'effet que « l'explication du raisonnement » par les élèves ne représente pas une activité en soi, tel que le précise la question.

Des commentaires ont également été compilés au sujet de la formulation du deuxième item de la TEIMS traitant de la fréquence à laquelle les enseignants demandent à leurs élèves d'écrire des équations et des fonctions pour représenter des relations. Deux des participants affirment que la formulation de la question TEIMS leur semble acceptable. Les deux autres ne répondent pas directement à cette question. Les panélistes s'entendent tous pour dire que la formulation respecte les règles reconnues en matière de syntaxes et de grammaire et que la longueur de la question leur semble tout à fait acceptable. Deux enseignants ajoutent cependant une note à l'effet que l'acronyme de la « TEIMS » leur soit inconnu. Il n'y a aucun commentaire négatif au sujet de la formulation de cette deuxième question.

Pour ce qui est de la clarté de la question, trois personnes sur quatre ont répondu que la question du premier item de la TEIMS n'était pas facilement comprise et que l'information demandée au répondant n'est pas clairement identifiée. Ces trois participants

poursuivent en précisant que la question pourrait facilement être interprétée de diverses façons. Toutefois, pour le quatrième répondant, la question est facilement comprise et l'information demandée est claire. Cette personne a interprété la question d'une seule façon mais laisse un doute quant aux différentes interprétations possibles. Le deuxième item présenté dans le cadre du sondage semble plus clair pour les participants, tous affirment que la question est facilement comprise et que l'information présentée est claire. Cependant, la moitié des participants du panel croient que la question de la TEIMS pourrait être interprétée de diverses façons, entre autres puisque les variables mises en relation ne sont pas précisées par la question. Les deux autres participants ont du mal à imaginer une interprétation différente de la leur et affirme que la question est bien dirigée. Toutefois, un panéliste ajoute que la clarté de cette question dépendra du niveau de connaissances mathématiques du répondant.

Les quatre participants du panel ont également répondu à des questions portant spécifiquement sur leur *niveau de familiarité avec le vocabulaire utilisé* dans les deux items de la TEIMS. Les résultats sont mitigés pour le premier item du sondage portant sur l'explication du raisonnement. Selon les réponses des participants, le vocabulaire utilisé dans la question semble tout à fait familier pour deux des participants et très peu familier pour le reste. Cependant, même ceux ayant rapporté une familiarité avec le vocabulaire de la question, précisent que certains mots, comme le mot « idée », ne possèdent pas une définition bien établie ou uniforme chez leurs collègues. Notamment, le terme « idée » ne serait pas suffisamment précis et le lien qu'il occupe avec le mot « raisonnement » dans cette question ne serait pas clair pour au moins un des répondants. Un participant a de la difficulté à saisir le sens du mot « sous-tend » et l'associe à l'expression « sous-entend ». À ce sujet, un des répondants précise que la formulation « sous-tend » ne serait pas utilisée dans le

cadre de son travail et un autre ajoute que ce vocabulaire est rarement utilisé avec ses collègues. Deux participants rapportent toutefois que le terme « raisonnement » est utilisé couramment dans leur milieu. Deux des répondants résumant la question en leurs propres mots et arrivent ainsi à démontrer deux interprétations semblables de la question.

Au niveau du second item, les résultats sont un peu plus uniformes, tous les participants affirment tous que les termes de la question leurs sont familiers et que leurs collègues connaissent et utilisent aussi ce langage. Toutefois, certains panélistes expliquent que même s'ils connaissent la signification du mot « fonction », ils ne l'utilisent pas dans l'enseignement des mathématiques pour les élèves de la 7^e ou 8^e année. Au sujet du sens accordé aux termes importants de la question, les membres du panel semblent bien connaître le sens des principaux vocables de la question (par exemple : équation, relation et fonction). Un participant note que la terminologie est importante et que ses collègues et lui tentent de l'uniformiser afin d'aider les élèves dans leur cheminement d'une année à une autre.

Présence de définition pour le vocabulaire utilisé - Suite à l'examen des versions officielles canadiennes françaises des questionnaires contextuels de la TEIMS (1995, 1999 et 2003), aucune ne comportent de lexique, de glossaire ou même de définitions des termes utilisés. Ceci s'applique notamment dans le cas des termes décrivant des pratiques prescrites par les récents courants de réforme.

Les résultats diffèrent chez les répondants du panel en ce qui concerne *la présence de plusieurs concepts dans les questions* de la TEIMS. Deux panélistes sur quatre considèrent que le premier item de la TEIMS cible un concept unique. Un des répondant rapporte plus d'un concept ciblé par la question et le dernier membre du panel explique simplement que les concepts ciblés par la question ne sont pas précisés. Les participants n'ont pas fourni de réponse pour le deuxième aspect de ce thème, soit la possibilité que la

présence de multiples concepts puisse agir sur la clarté de la question. Deux enseignants du panel indiquent que divers concepts sont ciblés par le deuxième item de la TEIMS. Un de ces participants poursuit son explication en précisant comment l'inclusion de plusieurs concepts ne lui permet pas de répondre de façon précise à la question. Cet individu enseigne les équations aux élèves de 7^e et 8^e année, mais non les fonctions. Par conséquent, il explique ne pas pouvoir répondre avec justesse à cette question, puisque la fréquence d'utilisation varie de façon importante pour les deux concepts ciblés par l'unique question. Parmi les participants, deux d'entre-eux croient que la présence de multiples concepts dans une même question pourrait nuire à la compréhension des répondants. Un autre participant affirme que la question de la TEIMS « va dans une seule direction » mais qu'en mathématiques les pratiques décrites par ces questions correspondent à un grand nombre d'applications variées. Selon cet individu, les questions TEIMS semblent cibler un seul concept à la fois.

Les participants du panel ont répondu à trois questions différentes en lien avec la *pertinence des items TEIMS pour le contexte de réforme francophone de l'Ontario*. La première question leur demandait de préciser d'où proviennent leurs connaissances en lien avec le sujet des questions TEIMS. En général, ce sont à partir de cours universitaires, de manuels scolaires, d'ateliers, de publications du ministère et d'expérience dans le milieu que les répondants ont acquis leurs connaissances en lien avec les concepts clés des questions TEIMS. Les questions du panel visaient également à savoir si le contenu des questions TEIMS semble important pour le contexte actuel franco-ontarien. En ce qui a trait au premier item TEIMS, traitant de l'explication fournie par les élèves au sujet de leur raisonnement, les opinions sont partagées. La moitié des membres du panel croient que le contenu de la question est important pour le contexte scolaire francophone de l'Ontario, et

expliquent comment cette question s'apparente bien aux objectifs poursuivis par le curriculum de 1997. Un participant a les propos suivants : « Selon moi, la nouvelle réforme encourage non seulement l'acquisition des concepts et l'application de ceux-ci mais également l'habileté de penser, de réfléchir et de faire des choix. » Cependant, la question semble avoir été interprétée différemment par les participants, i.e. que certains ont considéré le facteur linguistique comme étant dominant et d'autres ont davantage considéré le contexte scolaire tel que le curriculum en place. Les participants ayant répondu en fonction du curriculum ont fourni des réponses positives, tandis que les participants ayant interprété les questions plus au niveau du contexte linguistique ont répondu négativement. Trois participants expliquent qu'il existe certainement un lien entre le rendement des élèves et le contenu abordé par l'item TEIMS. Un d'entre-eux précise qu'afin d'atteindre les compétences fixées par le MEFO, les élèves doivent être en mesure d'expliquer leur raisonnement pour démontrer une compréhension approfondie de la matière. Cependant, le quatrième participant s'oppose à cette affirmation et considère que l'item ne propose aucune piste pour expliquer le rendement des élèves.

En ce qui concerne le deuxième item TEIMS, celui portant sur l'écriture d'équations et de fonctions pour représenter des relations, les opinions divergent de nouveau. Un répondant soutient entièrement l'idée que le contenu de la question soit pertinent pour le contexte scolaire des francophones en Ontario tandis que deux autres nient cette pertinence. Il semblerait encore une fois, que les personnes ayant considéré le contexte linguistique ont davantage répondu négativement. Trois participants sur quatre admettent la possibilité d'un lien entre le sujet de l'item TEIMS et le rendement des élèves. Un participant tente de démontrer la pertinence de l'item et son rôle envers le rendement des élèves par la phrase qui suit : « Les élèves sont évalués sur la représentation des fonctions dans le plan cartésien ainsi

que sur la résolution d'équations. ». Un autre individu précise que cela permet d'évaluer si l'élève comprend bien la relation observée.

De plus, les commentaires des enseignants du panel au sujet de la *pertinence des choix de réponse* ont été recueillis et analysés. Certains répondants affirment que les choix de réponse présentés sont tout à fait adéquats et qu'ils permettent de répondre aux questions de façon précise en traduisant leur réalité de façon fidèle. Cependant, un des individus propose des changements aux échelles de réponse. Pour le premier item, il suggère d'ajouter un choix au milieu des options présentées, soit « la moitié du temps ». Pour faire correspondre le premier choix de la liste au dernier, il propose également de modifier le dernier choix pour : « toujours ou presque toujours ». Dans les réponses au sujet du deuxième item de la TEIMS deux individus précisent une fois de plus des changements pour améliorer l'échelle de réponse fournie. Un participant suggère encore une fois d'ajouter une option à la deuxième position, soit « la plupart des cours » et d'adoucir les premier et dernier choix en ajoutant l'additif « ou presque ». Au sujet de ce deuxième item, un des participants explique aussi que les choix ne sont pas vraiment appropriés pour la question car il n'enseigne pas les relations toute l'année et il est donc difficile de répondre soit « à tous les cours » ou « la moitié des cours ». Les choix seraient davantage pertinents et précis si on précisait une courte période de référence comme par exemple un module ou un unité scolaire.

Afin d'alléger la tâche des répondants, la majorité des questionnaires contextuels d'évaluations à grande échelle contiennent surtout des items dont l'énoncé principal s'applique à un grand nombre de terminaisons ou de *sous-questions*. Les questionnaires de la TEIMS 1995, 1999 et 2003 comportent en effet un grand nombre de sous-questions. Presque tous les items sélectionnés dans le cadre de la présente étude sont des sous-questions faisant

suite à un énoncé général. Pour les cycles de 1995 et de 1999, 22 des 23 items sélectionnés pour l'étude sont des sous-questions, pour 2003 il s'agit de 24 items sur un total de 26. La position de la sous-question dans la série de terminaisons proposées peut avoir un effet sur la compréhension des participants. Le tableau 16 fournit une répartition des items retenus pour la présente étude en fonction de la position occupée par chacune des sous-questions. Par exemple, en 1995, 18% des items sélectionnés représentent une sous-question au second rang d'une énumération.

Tableau 16

Pourcentage d'items à chaque position occupée dans l'ordre des sous-questions

Cycle d'enquête	Rang de la sous-question								
	1 ^{er}	2 ^e	3 ^e	4 ^e	5 ^e	6 ^e	7 ^e	8 ^e	9 ^e
	%								
1995	9	18	14	14	23	9	5	9	-
1999	9	18	14	14	23	9	5	9	-
2003	8	13	25	21	8	4	13	4	4

Dans les questionnaires de la TEIMS, peu d'items affichent clairement *une période de référence*. Parmi les items sélectionnés pour cette étude, aucun ne comporte de période de référence fixe. Il devient alors impossible de déterminer si les questions de la TEIMS au sujet des pratiques et des croyances des enseignants se réfèrent à la dernière année scolaire, au dernier mois ou encore aux derniers cours enseignés. Plus d'une page d'instructions sont fournies au début du questionnaire mais aucune d'entre-elles ne comportent d'information sur la période de référence. Le questionnaire comprend des indications générales dans les

énoncés de questions du genre : « Pendant les cours de mathématiques,... », « Si vous donnez des devoirs de mathématiques... » ou « Lors de l'évaluation des travaux des élèves... ». Ces énoncés précisent à quel moment dans leurs cours de mathématiques les enseignants devraient se référer sans définir une limite dans le temps pour restreindre quels cours doivent être considérés.

Les analyses évaluant *la sensibilité des items* ont été réalisées avec les 17 items qui se répètent dans les questionnaires des trois cycles d'enquête. Dans la pratique, l'ANOVA est une procédure statistique très robuste (Howell, 1998). Cependant, elle peut être compromise si les conditions d'application sont enfreintes de façon extrême. D'abord, l'analyse de variance sous-entend une homogénéité des variances, i.e. que chacune des populations présente la même variance. Dans le cadre de la présente étude, des tests ont été réalisés dans le but d'évaluer si les données permettent de respecter les postulats de base. D'abord, une vérification sur chacun des items permet d'identifier si la plus grande variance serait 4 fois plus grande que la plus petite des variances, en plus de la statistique de Levene qui évalue autrement l'homogénéité des variances. Le test Levene vérifie l'hypothèse nulle qui suit, $H_0 : \sigma^2_1 = \sigma^2_2 = \sigma^2_3$. Si les résultats sont significatifs et que l'hypothèse nulle est rejetée, alors cela indique que les variances ne sont pas équivalentes pour les populations des différents cycles. Toutefois, l'ANOVA n'est pas trop sensible aux déviations en terme d'égalité des variances si les échantillons sont suffisamment grands et contiennent un nombre semblable de dossiers. Les données doivent également répondre à la condition de normalité « qui revient en réalité à dire que l'erreur est normalement distribuée à l'intérieur des conditions » (Howell, 1998, p.343). La distribution des données est évaluée selon les aspects suivants : mesures de tendance centrale, mesures de dispersion, asymétrie, kurtose et représentation graphique des fréquences. De plus, Howell (1998) stipule que les écarts

modérés par rapport à la normalité ne sont généralement pas trop graves, sans préciser exactement ce qu'il signifie par « écarts modérés ». La troisième condition pose l'indépendance des observations, où il est postulé que connaître une observation ne permet pas d'en prédire une autre au sein de la même population. Ce postulat est respecté par les données de l'étude.

L'analyse de variance réalisée avec le logiciel SPSS (2001) a permis d'identifier des différences statistiquement significatives ($p = 0,05$) entre les moyennes des cycles d'enquête pour 16 items. Chacune de ces différences significatives a été confirmée par les procédures robustes de Welch et de Brown-Forsythe ainsi que par les analyses non-paramétriques équivalentes soit le test de Kruskal-Wallis. Ces procédures sont généralement plus robustes et produisent de meilleurs résultats lorsque les données dérogent des postulats de base. Les résultats de la méthode traditionnelle ainsi que ceux des méthodes robustes sont rapportés dans la section des résultats.

Cette première phase d'analyse permet de repérer les items possédant des différences significatives entre les cycles d'enquête, mais ne peut identifier où se situent ces différences. Pour ce faire, des évaluations ANOVA *a posteriori* ont aussi été réalisées dans le but de connaître plus précisément la nature des différences significatives entre les échantillons de 1995, 1999 ou 2003. La méthode du « test de la plus petite différence significative de Fisher », ou LSD (Least Square Difference), a été utilisée avec un alpha de 0,05 pour identifier les différences entre les cycles puisque Howell (1998) affirme que : « Si votre expérience implique trois moyennes, la procédure LSD s'impose parce que EE (le taux d'erreur de l'ensemble) restera à α (alpha), et vous bénéficiez de l'augmentation de puissance liée à l'utilisation des test *t* standard ». Le tableau des résultats complets relatant

toutes les distinctions signalées par ces analyses et la taille de l'effet expérimental se trouve en annexe E.

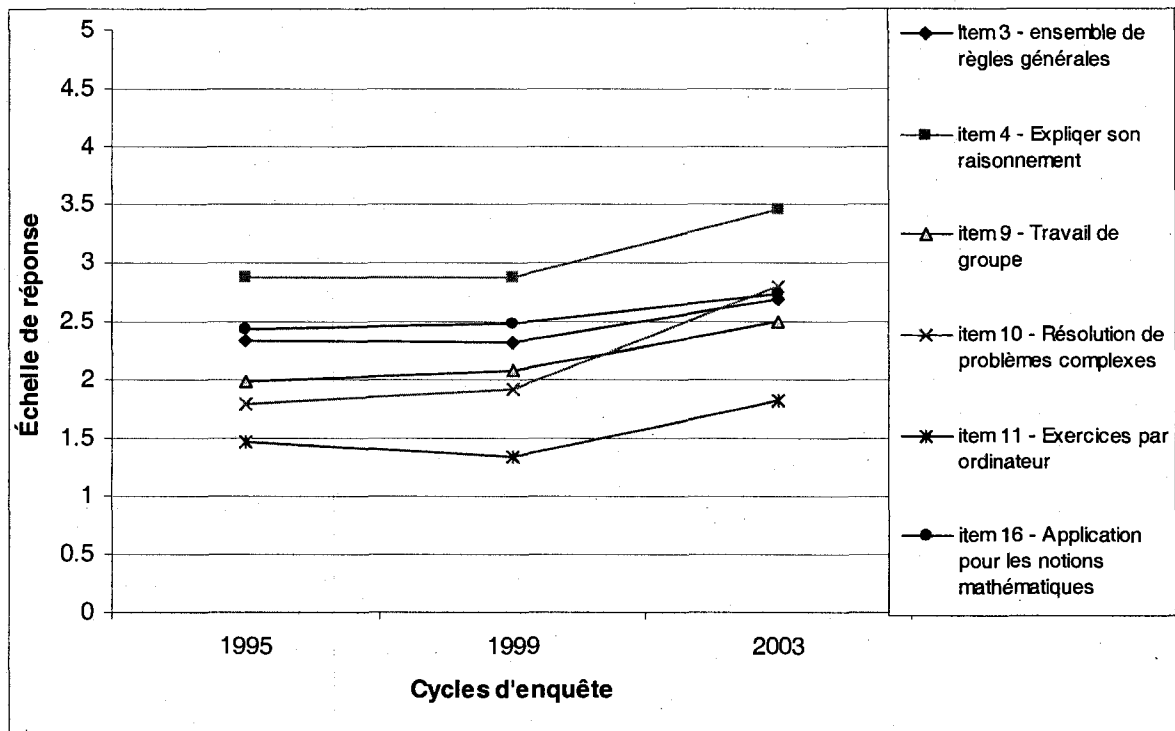


Figure 8. Courbes des moyennes d'items démontrant des hausses significatives entre les cycles

Afin de présenter clairement les résultats, des graphiques regroupent les items affichant des tendances semblables entre les cycles, i.e. la figure 8 contient uniquement les items dont les moyennes augmentent entre les cycles, la figure 9 représente les items dont les moyennes diminuent et la figure 10 contient les items dont les moyennes ne suivent pas de tendance continue entre les trois cycles d'enquête. Uniquement les items pour lesquels des différences statistiquement significatives ont été rapportées par les analyses sont présentés dans ces tableaux.

Des augmentations significatives sont notées entre 1995 et 2003 pour tous les items de la figure 8. Plusieurs de ces items ont des moyennes semblables et souvent non significatives entre les administrations de 1995 et de 1999, mais subissent des augmentations plus marquées en 2003 (items 3, 4, 9 et 16). L'item affichant les plus grandes moyennes porte sur l'explication du raisonnement par l'élève, il s'agit également du premier item présenté au panel d'enseignant pour recueillir leurs commentaires. Dans la figure 8, l'item ayant les plus faibles moyennes traite de l'utilisation de l'ordinateur pour effectuer des exercices ou résoudre des problèmes mathématiques. Les plus grandes augmentations caractérisent les items qui s'apparentent clairement aux recommandations pédagogiques instaurées par le nouveau curriculum de 1997 dont : l'explication du raisonnement, la résolution de problèmes complexes, le travail en petits groupes et l'utilisation d'ordinateurs. Toutefois, même si les différences observées entre les moyennes sont statistiquement significatives, la taille de l'effet (η^2) demeure faible pour la majorité des items. En effet, pour les items de la figure 8, la proportion de variance expliquée par la variable indépendante, soit les différents cycles d'enquête, se situent entre 1,7 % et 32,0 %. Il importe de souligner que η^2 équivaut à 0,320 pour l'item 10, un résultat qui suppose un effet important.

Cinq des six items illustrés portent sur des fréquences d'activités pédagogiques et possèdent des échelles de réponse identiques entre les items mais non entre les cycles (à quatre choix), tandis que l'item 3 se rapporte à une croyance pédagogique et utilise une échelle différente des autres items mais comporte également quatre choix de réponse. La formulation des questions demeure très semblable entre les cycles pour tous les items de la figure 8 et le nombre de choix de réponse fût modifié pour un seul item entre les cycles de 1999 et de 2003 (item 16).

Les items de la figure 9 affichent des baisses significatives de moyennes entre les cycles de 1995 et de 2003 ($p = 0,05$). L'item 5 se démarque par une moyenne très élevée en 1995 comparativement aux moyennes de 1999 et de 2003. En 1995, cet item comportait cinq choix de réponse, mais lors des enquêtes de 1999 et de 2003, les choix de réponse de l'item se sont vus réduits à trois options.

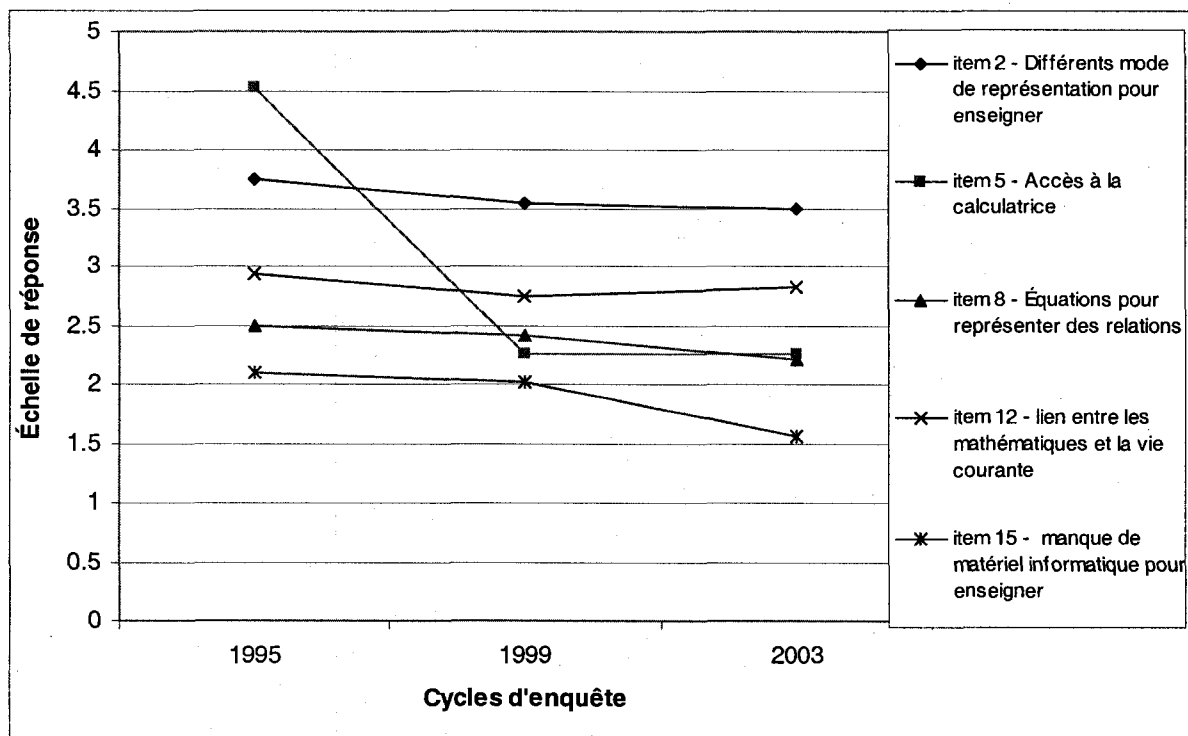


Figure 9. Courbes des moyennes d'items démontrant des baisses significatives entre les cycles

Dans la figure 9, l'item 2 affiche les plus grandes fréquences et l'item 15 les plus faibles. Les trois items ayant les fréquences les plus élevées traitent de sujets liés de près au contenu du curriculum de 1997 et ne présentent que de très faibles baisses entre 1995 et 2003. Deux des cinq items de cette figure portent sur des fréquences de pratiques pédagogiques et les trois autres reflètent des croyances en ligne avec les recommandations

du curriculum. La taille de l'effet (η^2) demeure très faible pour la majorité des items, soit entre 2,6 % et 5,0 %. En effet, seul l'item 5 pour lequel les changements d'échelle ont déjà été expliqués, affiche un pourcentage de variance expliqué élevé (67,3 %). La formulation des questions et les échelles de réponse comportent de légères différences entre les cycles pour les autres items de la figure 9.

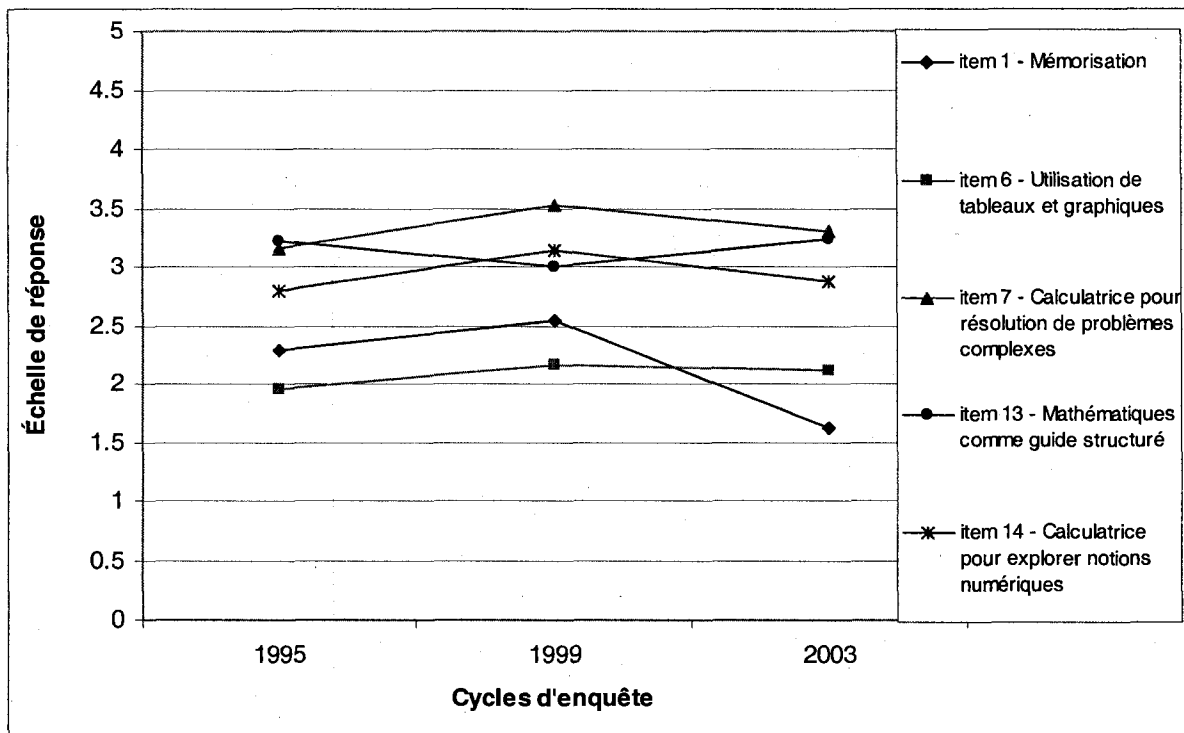


Figure 10. Courbes des moyennes d'items ne démontrant pas de tendance linéaire

Les items de la figure 10 affichent des différences significatives entre certaines de leurs moyennes mais aucune tendance linéaire uniforme à travers les cycles. Deux types de changements s'opèrent parmi ces items. Le premier modèle regroupe les items pour lesquels les moyennes étaient semblables en 1995 et en 2003 mais qui affichent une moyenne plus élevée en 1999 (items 6, 7 et 14). Le deuxième modèle présente, encore une fois, des moyennes très semblables pour 1995 et 2003 mais une moyenne significativement plus

faible en 1999 (item 13). L'item 1 fait exception à ces catégorisations et affiche une faible augmentation entre les administrations de 1995 et de 1999 et une très grande diminution en 2003. Cet item a subi des modifications au niveau du nombre de choix de réponse entre 1999 et 2003, i.e. il comportait trois choix de réponse en 1995 et en 1999, et quatre en 2003. Il fait appel aux croyances des enseignants sur l'importance de la mémorisation des concepts mathématiques et devrait afficher des fréquences plus faibles que d'autres items davantage liés aux recommandations du MEFO. Les items 7 et 14 sont issus du même énoncé de question, ils présentent tous les deux des fréquences élevées et des courbes très semblables à travers les cycles. Les items 1 et 13 portent sur des croyances tandis que les autres items illustrés dans la figure 10 traitent des fréquences des activités pédagogiques. De légères variations surviennent au niveau des questions et des choix de réponse pour l'ensemble des items. Toutefois, même si les différences observées entre les moyennes sont statistiquement significatives, la taille de l'effet (η^2) s'avère extrêmement faible pour la majorité des items. En effet, pour la majorité des items de la figure 10, la proportion de variance expliquée par la variable indépendante, se situent entre 2,2 % et 3,5 %. On remarque cependant que η^2 équivaut à 0,281 pour l'item 1, un résultat davantage explicatif.

Finalement, puisque la taille de l'effet représente une estimation de la corrélation maximale au carré entre la variable dépendante et la variable indépendante, il est normal que de faibles variations au niveau des moyennes entraînent des effets de petite taille. Il s'agit encore une fois de l'impact du type de variable générées (i.e. catégorielle) par les questionnaires contextuels sur les analyses statistiques.

Concordance entre élève et enseignant

Les résultats d'analyses rapportés dans cette section concernent la concordance pour certains items de l'étude qui s'effectue en comparant l'information rapportée par les enseignants à celle de leurs élèves répondant à des questions presque identiques. Cinq items ont été comparés entre les enseignants et leurs élèves pour les échantillons de 1995 et de 1999, tandis que l'échantillon de 2003 compte dix items comparés. Les données ont été transformées en cotes Z avant de réaliser les analyses de comparaison des scores. La procédure utilisée fût inspirée par les auteurs de deux rapports rédigés pour le NCES (1999, 2002) aux États-Unis et consiste à produire des scores de déviation moyenne pour les élèves d'une même classe et ensuite entre les élèves de la classe et leur enseignant respectif.

Le tableau 17 contient les résultats des déviations moyennes, rapportées en cotes z , entre les élèves pour les trois cycles. Les cases ombragées indiquent les items pour lesquels il est possible de faire une comparaison entre les cycles.

Tableau 17

Résultats standardisés des déviations moyennes entre les élèves d'une même classe

N° item	Contenu de la question	1995	1999	2003
Item 1	Croyances mathématiques (mémorisation)	0,7518	0,7745	
Item 4	Activité d'apprentissage (expliquer raisonnement)			0,6461
Item 5	Activité d'apprentissage (utilisation calculatrice)	0,5943	0,6487	0,7814
Item 6	Activité d'apprentissage (interprétation graphique)			0,7404
Item 8	Activité d'apprentissage (équations et fonctions)			0,7536
Item 9	Activité d'apprentissage (travail de groupe)	0,6447	0,6492	0,6355
Item 11	Activité d'apprentissage (utilisation de l'ordinateur)	0,6061	0,6447	
Item 12	Croyance mathématique (importance au quotidien)	0,7216	0,8062	0,7474
Item 26	Activité d'apprentissage (résolution de problèmes)			0,7629
Item 27	Activité d'apprentissage (problèmes complexes)			0,7911
Item 30	Activité d'apprentissage (TIC recherche d'information)			0,7877
Item 31	Activité d'apprentissage (TIC traitement et analyse)			0,7939

Les items pour lesquels la déviation moyenne (entre les élèves d'une même classe) se rapproche de zéro, comportent une grande similitude entre les réponses des élèves de chaque classe. Pour l'ensemble des variables, les résultats semblent indiquer des écarts comparables en fait de grandeur. Les coefficients de déviation varient entre 0,59 et 0,81. L'item 9, se démarque en affichant de faibles coefficients de déviation pour les trois cycles d'enquête (0,635 à 0,649). Cet item traite de la fréquence à laquelle les élèves travaillent en petits groupes pendant les cours de mathématiques. Des résultats semblables sont obtenus pour l'item 11 (0,606 à 0,645) qui porte sur l'utilisation des ordinateurs par les élèves pendant les

cours de mathématiques. L'item 1 présente des coefficients élevés (0,752 à 0,774) et concerne le degré d'importance accordée au processus de mémorisation pour la réussite mathématique. Les déviations moyennes obtenues pour l'item 12 s'avèrent également élevées (0,722 à 0,810), cet item rejoint le degré d'importance accordé par les élèves aux applications mathématiques dans leur vie quotidienne. En général, on remarque des déviations moyennes plus faibles, ce qui est désirable, pour les items qui traitent des activités d'apprentissage concrètes réalisées en classe comparativement aux items plus subjectifs se rapportant par exemple aux croyances mathématiques.

La seconde étape de l'analyse des déviations moyenne consiste à comparer les résultats moyens des élèves d'une classe donnée à ceux de leur enseignant (voir le tableau 18). De façon générale, ces déviations moyennes sont supérieures aux déviations observées uniquement entre les élèves d'une même classe, elles figurent donc entre 0,770 et 1,337. Les coefficients d'un même item varient souvent entre les cycles, notamment pour les items 5, 11 et 12. Les items affichant un plus grand coefficient démontre un plus faible taux d'accord entre l'information rapportée par l'enseignant et les propos de ses élèves.

Tableau 18

Résultats standardisés des déviations moyennes entre les élèves et leur enseignant

N° item	Contenu de la question	1995	1999	2003
Item 1	Croyances mathématiques (mémorisation)	1,178	1,192	
Item 4	Activité d'apprentissage (expliquer réponse)			0,940
Item 5	Activité d'apprentissage (utilisation calculatrice)	1,027	1,320	1,076
Item 6	Activité d'apprentissage (interprétation graphique)			1,144
Item 8	Activité d'apprentissage (équations et fonctions)			1,066
Item 9	Activité d'apprentissage (travail de groupe)	1,049	1,146	1,053
Item 11	Activité d'apprentissage (utilisation de l'ordinateur)	0,939	1,337	
Item 12	Croyance mathématique (importance au quotidien)	0,770	1,190	1,091
Item 26	Activité d'apprentissage (résolution de problèmes)			1,168
Item 27	Activité d'apprentissage (problèmes complexes)			1,135
Item 30	Activité d'apprentissage (TIC recherche d'information)			0,915
Item 31	Activité d'apprentissage (TIC traitement et analyse)			1,042

Validité théorique

Les analyses factorielles exploratoires (AFE) ont été réalisées afin de déterminer dans quelle mesure les items de la TEIMS se regroupent selon les concepts théoriques visés par l'étude. Différents facteurs ont limité l'utilité ou la valeur des analyses factorielles, dont les propriétés métriques des variables et la faible corrélation entre les items du questionnaire. En somme, deux phases d'analyses exploratoires ont été effectuées pour les données de chacun des trois cycles d'enquêtes. Les valeurs propres (*eigenvalues*) des modèles indiquaient en général des solutions avec un nombre très élevé de facteurs, laissant présager

l'instabilité des modèles. Une observation des matrices de corrélation a permis de voir qu'elles contiennent majoritairement des valeurs faibles et très faibles (inférieures à 0,3). Les estimations de la variance de l'erreur des variables produites pour chacune des solutions factorielles, se sont avérées élevées (souvent entre 0,7 et 0,9). Les indicateurs RMSR sont demeurés trop élevés, même en considérant des solutions factorielles de cinq ou de six facteurs. Les valeurs du RMSEA étaient raisonnables à partir des solutions à 4 facteurs mais ces solutions demeurent difficiles à interpréter et l'obtention d'une structure simple tant convoitée pour les analyses factorielles n'apparaissait pas de façon évidente. En général, pour les modèles en annexe F, les items semblent se regrouper selon leur forme et thème central, autant sinon plus que selon leur véritable contenu. C'est-à-dire que les sous-questions d'un même énoncé principal se retrouvent dans un même facteur, indiquant que les enseignants auraient davantage tendance à fournir des réponses semblables pour des items qui se rapprochent en format plutôt qu'en sens. Ceci peut être interprété comme un signe de contamination entre les questions ou encore de persévérance dans la façon de répondre.

Biais de réponse

Pour les questionnaires contextuels des enseignants, le phénomène de désirabilité sociale peut provoquer chez certains un biais systématique. Tous les membres du panel ont affirmé qu'ils ne ressentent aucune pression sociale pour répondre aux items tirés du questionnaire de la TEIMS. Ils affirment qu'ils se sentent tous confortables de répondre en toute franchise. Un des individus ajoute toutefois qu'un enseignant par exemple qui ne ferait jamais appel au raisonnement de ses élèves serait sûrement tenté de fournir une réponse qui épouse davantage les recommandations du curriculum.

Fidélité

Dans le cadre de cette étude, la fidélité des items de la TEIMS est évaluée selon la méthode des covariances. À l'aide du coefficient *alpha* de Cronbach, une estimation de la proportion de variance due au score vrai est obtenue, ce qui est communément appelé la cohérence ou fidélité interne. Les analyses de cohérence interne ont été lancées à partir de premiers regroupements basés sur une analyse de contenu théorique des items. Les items ont été regroupés en fonction du concept qu'ils étaient soupçonnés de mesurer, soit : le raisonnement mathématique, la résolution de problème ou la communication mathématique.

Les résultats de l'analyse, soit les coefficients alpha qui figurent dans le tableau 19 se situent entre 0,3315 et 0,6348. Les coefficients issus des analyses réalisées avec tous les items représentant les croyances et les pratiques pédagogiques des réformes sont supérieurs à ceux produits pour les trois catégorisations. Parmi ces dernières, la catégorie raisonnement/compréhension acquière des coefficients supérieurs en 1995 et 1999 bien que cela résulte probablement du plus grand nombre d'items. Toutefois, ce sont les variables mesurant la résolution de problème qui produisent le plus grand coefficient pour l'échantillon de 2003. En somme, les coefficients alpha obtenus pour les trois catégories sont relativement faibles, et cela laisse croire que les items de ces échelles ne seraient pas des mesures adéquates de ces thèmes.

Tableau 19

Résultats de l'analyse de fidélité des indices composés pour les trois thèmes de l'étude

Thèmes	Nombre d'items	Alpha de Cronbach (α)	Corrélation entre l'item et le score total	
			Minimum	Maximum
1995				
Raisonnement/compréhension	20	0,5529	0,0304	0,4521
Résolution de problèmes	11	0,3315	0,0201	0,3182
Communication mathématique	8	0,3875	0,0298	0,3947
Tous les items – pédagogie réforme	23	0,6172	0,0239	0,4104
1999				
Raisonnement/compréhension	20	0,6195	0,0265	0,4651
Résolution de problèmes	11	0,4611	0,0204	0,3503
Communication mathématique	8	0,3870	0,0867	0,3746
Tous les items – pédagogie réforme	23	0,6348	0,0334	0,4861
2003				
Raisonnement/compréhension	19	0,4724	0,0066	0,4195
Résolution de problèmes	15	0,5051	0,0492	0,5064
Communication mathématique	8	0,3571	0,0108	0,3292
Tous les items – pédagogie réforme	25	0,6074	0,0066	0,4195

Discussion

L'étude a comme objectif principal de répondre à la question de recherche suivante :
Quelle est la qualité psychométrique des items issus des questionnaires contextuels des enseignants de la TEIMS portant sur les pratiques d'enseignement (curriculum appliqué) ciblées par les réformes scolaires (curriculum projeté) ? Cette section interprète les résultats de l'étude de manière à pouvoir répondre à la question. Il traite de leur contribution au domaine, des interrogations qui demeurent et des limites de l'étude.

Sommaire des produits de la recherche

D'abord, les propos des enseignants du panel ont démontré que les deux items exemplaires de la TEIMS-R comportent des lacunes pouvant nuire à la compréhension des répondants. En outre, les panélistes rapportent des problèmes au niveau de la formulation et de la clarté des questions ainsi que du vocabulaire utilisé. Par exemple, un terme communément utilisé par les prétendants des réformes scolaires, le mot « raisonnement » a engendré des problèmes de compréhension chez les panélistes. Afin de pallier aux problématiques rencontrées dans le cadre de cette étude et dans d'autres recherches semblables (NCES, 1999 ; NCES, 2002 ; Porter & Smithson 2001), les items des questionnaires contextuels devraient être formulés de façon simple à l'aide d'un vocabulaire familier et pointu afin de permettre une meilleure compréhension et une interprétation uniforme chez les participants. Et, des exercices de validation des questions devraient être

réalisés davantage sur les questionnaires dont les résultats sont utilisés pour s'informer au sujet du fonctionnement du système scolaire (Porter & Smithson 2001).

Certains auteurs expliquent que si les termes utilisés dans les questionnaires y étaient clairement définis, cela pourrait améliorer la compréhension des participants (Ball & Rowan, 2004; Mullens & Kasprzyk, 1999; NCES, 1999, 2002; Porter & Smithson, 2001). La présente étude a permis de confirmer que les questionnaires de la TEIMS-R, comme bien d'autres du même genre, ne contiennent pas de définitions pour les termes utilisés. Cependant, l'ajout d'un lexique ou de définitions à même le questionnaire ne pourrait qu'ajouter au fardeau des répondants et contribuer davantage à la dégradation du taux de réponse et à la qualité de l'information fournie par les participants.

Des auteurs ont déjà souligné les problèmes de compréhension dus à des questions trop complexes (Goldstein, 1995; NCES, 2002). Ils exposent notamment que les questions comportant trop d'information doivent être lues, comprises, interprétées correctement et retenues dans la mémoire du répondant pour lui permettre de formuler une réponse adéquate. Ce processus cognitif devient mentalement complexe lorsque la question contient un grand nombre de concepts. Cette situation risque d'engendrer des erreurs ou le découragement chez les participants. Selon certains panélistes, la présence de plusieurs concepts dans une même question nuit à la compréhension des répondants et donc à la qualité psychométrique de l'item. Les résultats démontrent aussi que la présence de différents concepts dans une même question pourrait rendre l'item très peu fiable, l'exemple rapporté par un panéliste est présenté dans la section des résultats. La problématique soulevée concerne les items regroupant plusieurs concepts comme le suivant : « Pendant les cours de mathématiques, vous arrive-t-il de demander aux élèves de réaliser les activités suivantes ? Présenter et analyser des rapports à l'aide de tableaux, de diagrammes ou de graphiques ». Devant cet

item, on se doit de demander si les répondants additionnent les fois où ils utilisent ces médiums, ou bien s'ils font une moyenne de l'utilisation des trois, ou encore s'ils choisissent celui qui se produit le plus souvent et répondent uniquement en fonction de celui-là ?

En effet, certains énoncés sont fort simples et ne contiennent pas plus d'un ou deux éléments de contexte. À titre d'exemple, il s'agit de considérer l'item suivant tiré du questionnaire des enseignants de 1995 : « Dans quelle mesure êtes-vous d'accord avec chacun des énoncés suivants ? ». Par contre, d'autres énoncés de question comportent au moins trois éléments de contexte, que le répondant doit conserver dans sa mémoire active pendant qu'il s'affaire à répondre aux nombreuses sous-questions s'y rapportant. Les items retenus pour la présente étude sont associés à des énoncés de question de complexités diverses.

La compréhension de la question et la validité des réponses engendrées par celle-ci dépendent aussi de la pertinence du sujet abordé. Les participants du panel considèrent les items exemplaires de la TEIMS comme étant relativement pertinents pour leur contexte particulier d'enseignant de mathématiques des écoles françaises en Ontario. Les panélistes jugent que les thèmes ciblés par les questions de la TEIMS peuvent influencer le rendement de leurs élèves et ces thèmes sont directement abordés par le curriculum de 1997.

Si le contenu de la question doit être pertinent pour le répondant et son contexte, alors il en est de même pour la période de référence des items. La recension des écrits rappelle qu'afin d'assurer la qualité psychométrique d'un instrument, la période de référence doit au moins être identifiée clairement dans chaque item (Mayer, 1999; Rowan, Correnti & Miller, 2002). Peu d'items du questionnaire contextuel à l'intention des enseignants participants à la TEIMS affichent clairement une période de référence. Parmi les items sélectionnés pour cette étude, aucun ne comporte de période de référence fixe. Lorsque la

période de référence n'est pas indiquée, les enseignants répondent en fonction de la période qui leur convient, soit la dernière année scolaire, la présente année scolaire, le dernier mois, les dernières deux semaines ou encore les derniers cours. Dans une étude ayant examiné des items semblables à ceux de la TEIMS, où la période de référence n'est pas clairement indiquée, la majorité des enseignants ont utilisé l'année scolaire comme unité de référence (NCES, 2002). Cependant, il est important de noter que cette étude ainsi que celle de Porter et Smithson (2001) rapportent des problèmes de rappel chez les répondants pour une période aussi longue; un problème affligeant vraisemblablement aussi les données de la TEIMS.

Les analyses de variance ont produit des résultats intéressants mais laissent cependant planer un doute quant à la sensibilité des items de questionnaire contextuel. D'abord, les croyances et les pratiques des enseignants en matière de communication mathématique semblent concorder avec les recommandations émises par le MEFO, c'est-à-dire que les moyennes de fréquence pour ces items augmentent entre 1995 et 2003. Les enseignants de 2003 ont également affirmé que leurs élèves travaillent plus souvent avec les ordinateurs pour mettre en pratique les mathématiques comparativement aux enseignants de 1995 et de 1999, ce qui correspond encore une fois au contenu du curriculum de l'Ontario de 1997. Par ailleurs, les items clairement liés à la justification du raisonnement par les élèves démontrent une équivalence statistique entre les cycles de 1995 et de 1999 et une augmentation marquée en 2003. De plus, le curriculum de l'Ontario de 1997 offre un nombre imposant de recommandations incitant les enseignants à joindre l'apprentissage des mathématiques à des situations réelles de la vie. Parmi quatre items de l'étude portant sur le lien entre l'apprentissage des mathématiques et les situations de la vie réelle, deux affichent des moyennes stables entre 1995 et 2003 et deux autres démontrent de légères augmentations. Enfin, les items représentant expressément des activités de résolution de

problèmes, ont démontré un résultat stable entre 1995 et 1999 et une augmentation nette des fréquences moyennes entre 1995 et 2003.

Il est difficile d'expliquer les résultats obtenus pour un ensemble d'items ciblant tous des croyances et des pratiques recommandées par un même curriculum. Les différences de moyennes observées représentent soit de véritables tendances au sein de la population enseignante ou des problèmes de qualité psychométrique au niveau des items. Certes, à la lumière des différences observées parmi les caractéristiques des échantillons, il y avait lieu d'envisager des variations significatives à la hausse pour les items de l'étude. En outre, l'échantillon de 2003 contient un nombre important de nouveaux enseignants ayant reçu une formation récente. Si les programmes de formation des maîtres tiennent compte des nouvelles recommandations en vigueur dans la province de l'Ontario, cela devrait influencer les résultats rapportés par les enseignants de 2003 au sujet de leurs pratiques.

Peu d'études ont pu comparer les pratiques pédagogiques à travers le temps dans un contexte de renouveau pédagogique comme le fait la présente. Les mesures prises lors des cycles de la TEIMS correspondent à des moments cruciaux de l'implantation d'un nouveau curriculum dans la province de l'Ontario, soit avant (1995), pendant (1999) et après (2003) la venue du nouveau programme. Même si des recherches semblables n'ont pas été répertoriées, une étude réalisée par Porter et Smithson (2001), explique qu'au début des années 1990, les enseignants mettaient l'accent davantage sur la pratique et les connaissances procédurales, tandis que de nos jours, ils se concentrent plus souvent sur des tâches offrant un véritable défi cognitif pour les élèves telles que la résolution de problèmes complexes ou l'application des mathématiques dans le contexte de la vie courante. Les divers problèmes de qualité psychométrique constatés par les analyses réalisées dans le cadre de cette étude ont fort probablement contribué aux résultats peu concluants des analyses

ANOVA. Les comparaisons se compliquent également en raison des changements d'échelles ou de questions survenus entre les administrations cycliques du questionnaire.

Les résultats de l'analyse vérifiant la concordance entre élève et enseignant montrent des variations importantes entre les résultats des élèves et ceux de leur enseignant de mathématiques. Ces résultats vont dans le même sens que ceux obtenus par deux recherches menées aux États-Unis (NCES, 1999, 2002) sur les items du National Assessment of Educational Progress (NAEP). Il est difficile de mettre en contexte la taille des déviations observées à cause de différences méthodologiques entre ces deux recherches. Tout de même, il semble que les meilleurs taux d'accord entre les enseignants et leurs élèves apparaissent pour les items reflétant surtout des pratiques pédagogiques simples et facilement comprises par tous. Ce processus offre un potentiel intéressant pour les études de validation réalisée post-hoc, mais afin de maximiser son pouvoir, il faut considérer les trois aspects suivants avant de l'utiliser : 1) des items (questions et échelles) très semblables doivent être d'abord administrés aux enseignants et à leurs élèves, 2) ces items doivent porter sur des pratiques appliquées de façon uniforme pour tous les élèves d'une classe, et 3) les élèves participants doivent être suffisamment âgés car la réponse aux items est une opération cognitive exigeante. Comme pour la présente étude, Levine et ses collègues, ont obtenu des variations non négligeables entre les résultats des enseignants et ceux de leurs élèves (NCES, 2002). Ils attribuent ces différences aux facteurs suivants : une interprétation différente des questions par les participants, un manque d'habiletés cognitives nécessaires afin de synthétiser la fréquence des activités pédagogiques adéquatement (notamment chez les étudiants), surtout lorsque les activités se produisent rarement ou si la période de référence est ambiguë, trop longue, ou non indiquée (NCES, 2002).

L'examen de la validité théorique effectué par l'application d'analyses factorielles présente certains aspects positifs mais dans l'ensemble demeure peu convaincant. En effet, les items semblent se regrouper selon la forme et le thème général, plus que selon leur véritable contenu, plus subtil au premier regard. Idéalement, les résultats des analyses factorielles auraient démontré une structure comparable aux regroupements effectués en fonction des contenus théoriques présentés dans divers documents de réformes scolaires (MEFO, 1997; NCTM, 2000). Cela aurait contribué à valider ces items en tant que mesures adéquates des concepts théoriques sous-jacents. Certaines considérations statistiques pourraient aider à expliquer la médiocrité des résultats. D'abord, il est important que les variables incluses dans ce type d'analyse aient un minimum de variance. Lorsque des analyses factorielles sont réalisées à partir de données catégorielles, la variance de chaque item est restreinte. Ensuite, les variables devraient être distribuées approximativement normalement (surtout pour des analyses factorielles confirmatoires). De plus, les variables doivent être suffisamment corrélées entre elles pour permettre la formation d'un modèle simple qui représente bien les données empiriques.

Les analyses réalisées dans le but d'établir les propriétés métriques des données de la TEIMS ont servi à discerner certaines problématiques inquiétantes liées à la nature des données. D'abord, les données générées par les questionnaires contextuels des enseignants proviennent presque uniquement de variables catégorielles. Ensuite, les distributions des items ont révélé plusieurs dérogations par rapport à une distribution normale. Tous les items de l'étude qui portent sur l'utilisation d'ordinateurs pour l'enseignement ou l'apprentissage des mathématiques ont des fréquences très faibles et souvent des choix de réponse ignorés dans l'extrémité supérieure de l'échelle de réponse. Ces résultats sont identiques à d'autres obtenus lors d'une étude réalisée avec les données du NAEP (NCES, 1999). Plusieurs items

au sujet des croyances pédagogiques ont également des distributions problématiques. Il s'agit principalement d'items tels que : « Pour qu'ils réussissent en mathématiques à l'école, dans quelle mesure pensez-vous qu'il est important pour les élèves de ... » ou « Dans quelle mesure êtes-vous d'accord avec chacun des énoncés suivants ? ». L'examen des mesures de tendances centrales, des mesures de dispersion, de l'asymétrie, de la kurtose et des tests de distribution normale permettent de conclure que les caractéristiques distributives de ces variables limitent sérieusement l'utilisation que les analystes peuvent en faire. Cependant, dans bien des situations, ces données ont été, et seront de nouveau, utilisées sans tenir compte des effets pervers que ces caractéristiques peuvent entraîner au niveau des résultats d'analyses statistiques.

Il est important de considérer la pertinence des choix de réponses lorsque la distribution d'un item s'avère problématique. Les commentaires des panélistes au sujet de la pertinence des choix de réponse sont généralement positifs. Toutefois, ils suggèrent d'ajouter un choix de réponse afin d'équilibrer l'échelle de réponse. Une échelle à cinq choix semble préférable selon les répondants afin de permettre une plus grande variété de réponses. Les panélistes expliquent l'importance de conserver des termes équivalents aux deux extrémités de l'échelle de réponse.

La pertinence du contenu des items a un impact sur les problèmes liés à la distribution des réponses. Par exemple, des fréquences beaucoup trop élevées suggèrent peut-être des concepts trop généraux ou communs chez l'ensemble des enseignants et dans ce cas les contenus devraient être précisés davantage. Les choix de réponse devraient également faire l'objet d'une révision afin de s'assurer qu'ils permettent de couvrir toutes les possibilités de réponse et de produire une variance potable au niveau de l'item. En effet, un item désirable serait un item qui produirait une quantité maximale d'information au sujet des

répondants, c'est-à-dire un item qui possède une grande variance et une distribution près de la normale. Sans ces attributs importants, le potentiel de l'item en tant que « générateur d'information » se trouve grandement limité.

Les commentaires des panélistes portent à croire que le phénomène de désirabilité sociale ne provoque pas de biais systématique pour les items de l'étude. D'autres études ont démontré la présence d'un tel biais surtout lorsqu'il s'agit d'items portant sur des pratiques pédagogiques de réformes mais signalent également que l'effet d'un tel biais serait presque imperceptible (Burstein et al. 1995; Mayer, 1999; NCES, 1999, NCES, 2002; Smithson & Porter, 2001). Les items devraient tout de même toujours être formulés de façon neutre afin de ne pas encourager un tel biais de réponse. Des recherches futures devraient être élaborées afin que l'information rapportée par les enseignants au sujet de leurs pratiques puissent être comparées aux jugements d'observateurs neutres, permettant ainsi de déterminer avec plus de précision si les enseignants sont tentés de rapporter une plus grande utilisation des pratiques de réformes.

La fidélité interne des items, évaluée selon la méthode des covariances, a produit des résultats encore une fois assez peu convaincants. En effet, les échelles créées selon les concepts pédagogiques issus des réformes (raisonnement, résolution de problèmes et communication) ne semblent pas faire preuve d'une grande cohésion puisque les matrices de covariances affichaient des valeurs faibles. Il faudra des analyses subséquentes afin de déterminer quels items pourraient être regroupés afin de mesurer adéquatement les concepts importants des réformes scolaires actuellement en vigueur dans les écoles du pays.

Contextualisation des résultats de l'étude

Les analyses élaborées dans le cadre de la présente recherche avaient comme objectif principal d'évaluer la qualité psychométrique de certains items du questionnaire contextuel de la TEIMS-R, à l'intention des enseignants. Ces items portent sur les croyances et les pratiques pédagogiques faisant l'objet des plus récentes réformes scolaires. La plupart de ces efforts de réforme s'inspirent des recommandations émises par le NCTM qui préconise un enseignement axé sur le raisonnement et la compréhension, la résolution de problèmes et la communication mathématiques (Grouwns & Cebulla, 2000). Les réformateurs tentent d'améliorer les systèmes d'éducation, voire la qualité de l'enseignement dispensé, en instaurant des recommandations surtout au niveau des pratiques pédagogiques (Mayer, 1999). Bien que les recommandations de la NCTM et du MEFO possèdent plusieurs similitudes et visent à établir de meilleures conditions d'apprentissage, le programme de l'Ontario demeure vague quant aux pratiques pédagogiques précises à appliquer en salle de classe. Conséquemment, les enseignants jouissent d'une grande liberté en ce qui concerne le choix de leurs pratiques pédagogiques. Il devient donc difficile de savoir dans quelle mesure cette ouverture a un effet positif sur les apprentissages et incite les enseignants à varier leurs pratiques et à innover dans ce sens. Les recherches réalisées à ce jour dans des contextes de réforme démontrent plutôt le contraire, c'est-à-dire qu'en absence de modèles clairs, les enseignants ont souvent le réflexe d'adapter légèrement leurs pratiques usuelles (Stecher et al., 2002). Ainsi, chaque enseignant crée des pratiques « hybrides » qui varient énormément d'un enseignant à un autre. Les enseignants utilisent ensuite la nomenclature des pratiques introduites par les réformes pour décrire ou nommer ces pédagogies hybrides. En bout de ligne, les termes conçus pour désigner des pratiques novatrices précises sont utilisés pour

décrire une grande variété de pratiques divergentes créées autour des thèmes introduits par les réformateurs. Il devient alors difficile d'établir une définition uniforme des stratégies pédagogiques pratiquées actuellement en salle de classe.

Les efforts de réforme déployés de nos jours surpassent toutes les démarches semblables réalisées par le passé (Burstein et al., 1995). Afin de pouvoir guider leur mise en œuvre, une information juste et valide décrivant ce qui se passe dans les salles de classes s'avère nécessaire. Jusqu'à maintenant, les données disponibles au sujet des pratiques pédagogiques sont en quantité limitée et d'une qualité discutable (Burstein et al., 1995). Dans ce contexte, les données des évaluations à grande échelle, comme celles utilisées dans le cadre de la présente étude, prennent beaucoup d'importance. Plusieurs chercheurs (Crocker, 2002; Robitaille & Taylor, 2002) attribuent à ces programmes d'évaluation un réel potentiel et croient qu'ils peuvent contribuer de façon importante à l'amélioration des systèmes d'éducation canadiens. En effet, selon Earl et Torrance (2000) il y aurait consensus parmi les chercheurs du domaine à l'effet que les évaluations à grande échelle devraient exercer un impact positif sur les systèmes d'éducation et guider les modifications à effectuer. C'est sans doute pourquoi, les données reliées aux croyances et aux pratiques pédagogiques des enseignants, même celles recueillies à l'aide de questionnaires contextuels, suscitent tant d'intérêt.

En effet, selon la perspective des politiques éducatives, les enseignants sont les principaux artisans de la mise en œuvre des réformes, ils décident ultimement du contenu à enseigner et de la façon dont le contenu sera présenté aux élèves (NCES, 1999). De plus, plusieurs des facteurs déterminants du rendement des élèves ne peuvent être influencés par les politiques éducatives, comme les caractéristiques socio-économiques ou les habitudes de lecture des élèves à la maison. Donc, pour les réformateurs, les chercheurs, les

administrateurs et les directions scolaires, ce sont les données permettant d'expliquer les niveaux de rendement des élèves en fonction de facteurs liés à la classe ou à l'école qui portent le plus d'intérêt.

Les données au sujet des pratiques pédagogiques ont une grande valeur surtout lorsque l'on considère qu'en certains cas ils arrivent à créer un véritable impact sur la qualité de l'enseignement. Earl et Torrance (2000), proposent à titre d'exemple positif le cas des examens provinciaux de l'Ontario menés par l'EQAO. Leur étude rapporte que 90 % des écoles affirment avoir fait des changements aux pratiques pédagogiques suite aux recommandations émises en fonction des résultats obtenus aux tests. Cet approvisionnement viendrait combler un besoin important puisque selon Clare (2000) les enseignants ne possèdent pas toujours les outils ou les connaissances leur permettant d'améliorer la qualité de l'enseignement en fonction des lignes directrices du NCTM ou autres cadre de réforme.

Cependant, si les interprétations, conclusions et décisions forgées par les programmes d'évaluation ne sont pas valides, alors ces programmes deviennent davantage problématiques qu'utiles (Earl & Torrance, 2000). En effet, la présente étude rejoint les préoccupations de plusieurs auteurs qui émettent des mises en garde au sujet de la validité et de la fidélité des items de questionnaires contextuels et suggèrent des efforts additionnels afin de mieux évaluer la qualité psychométrique des ces items (Burstein, 1995; Mayer, 1999; NCES, 1999; NCES, 2002; Rowan et al., 2002; Smithson & Porter, 2001). Il n'est pas simple de porter un jugement final quant à la véritable utilité des données de questionnaires contextuels pour informer les décideurs en quête de données. En regardant les résultats de la présente étude, on constate que l'information générée par les questionnaires est limitée en terme de quantité et de qualité. D'abord, les limites évidentes en terme de « quantité » d'information pourraient s'expliquer par la production de variables presque exclusivement

catégorielles et des lacunes distributives qu'on leur associe. L'examen des propriétés métriques des items a permis de déceler un manque de « quantité » d'information produite par les items au niveau de la variance très limitée qu'ils comportent. Ce manque d'information mine sérieusement toute tentative d'analyse à base corrélationnelle.

D'ailleurs, les analyses de ce genre exécutées dans le cadre de la présente recherche ont produit des résultats de faible intensité ou de moindre signification. Les échelles utilisées pour rapporter les résultats des enseignants par rapport à certaines questions valables ne permettent pas de recueillir une variabilité de réponse suffisamment grande pour déceler les changements au niveau des pratiques des enseignants. L'administration coûteuse de ces questionnaires produit essentiellement une information à gros grain qui ne laisse point entrevoir les tendances pédagogiques émergentes. De plus, le manque de précision dans les données empêche probablement d'établir la relation avec la variable ultime, soit le rendement des élèves. Des résultats comportant davantage de finesse au niveau de l'information produite permettraient aux chercheurs de déceler les réalités pédagogiques avant qu'elles ne deviennent tellement répandues et évidentes que leur démonstration empirique n'apprend rien de nouveau aux individus concernés. Certes, l'information actuelle ne génère pas le niveau précision requis par les décideurs leur permettant de réagir à temps et d'élaborer les politiques éducatives adéquates, i.e. avant d'avoir à instaurer des correctifs exigeant aux enseignants d'effectuer un virage à 180 degrés. C'est un peu comme un navire naviguant dans un épais brouillard sans instruments de navigation. Au moment où l'iceberg est visible par les gens sur le pont du navire, il est déjà trop tard. À cette ère technologique, les commandants du monde de l'éducation méritent de meilleurs instruments de navigation, sans quoi, ils ne pourront être tenus responsables de la suite des évènements. De nombreux articles et ouvrages portent sur l'imputabilité des organisations face à l'apprentissage des

élèves. Toutefois, tout comme il serait malsain d'exiger qu'un élève effectue des problèmes de géométrie avec un instrument de mesure qui ne possède pas un degré de précision adéquat, il n'est pas juste de supposer que les dirigeants politiques pourront mener à bien nos systèmes d'éducation avec les données des questionnaires contextuels.

Par ailleurs, ces données sont également limitées en terme de « qualité » d'information. D'abord, la mesure des pratiques ou croyances pédagogiques devrait reposer avant tout sur une définition claire et précise des concepts à mesurer. Parmi les exemplaires consultés de devis d'élaboration émis par les concepteurs des questionnaires aucun ne comportaient de cadre théorique suffisamment solide et cohérent pour représenter le contenu des questionnaires contextuels. Les documents de référence décrivant l'organisation et le contenu des questionnaires, ressemblaient plutôt à un amalgame effectué à l'emporte pièce de plusieurs concepts reliés, de loin ou de près, à l'enseignant. Les croyances et les pratiques pédagogiques représentaient des catégories au sein d'organigrammes complexes dans lesquels plusieurs concepts importants étaient mis en relation de cause à effet. Si l'objectif des questionnaires est de mesurer un phénomène avec exactitude et précision, la première étape devrait consister en l'élaboration d'une définition claire et précise du domaine à mesurer. Les croyances et les pratiques pédagogiques n'ont pas été clairement définies par les concepteurs de questionnaires, par conséquent, l'utilisation des données à grande échelle est ardue pour les chercheurs qui n'ont d'autre choix que de tenter de bâtir des protocoles de recherche autour du contenu rapiécé des questionnaires contextuels. Non seulement le manque de précision discuté précédemment empêche-t-il les chercheurs d'arriver à des conclusions réellement significatives, mais une fois les résultats connus, ces derniers doivent également tenter de donner un véritable sens à ces données qui peuvent parfois paraître

décousues à cause d'un manque de cohérence ou de couverture dans le contenu des questionnaires.

La « qualité » d'information générée par les questionnaires est limitée non seulement en terme d'ampleur ou de cohérence des contenus mais la profondeur de l'information y est tout aussi défailante. Ce concept n'a pas été abordé au cours de la présente étude mais mérite d'être discuté brièvement puisqu'il vient influencer l'utilité des questionnaires. Tel que mentionné précédemment, les chercheurs sont limités par les sujets ou contenus sélectionnés lors de la conception des évaluations à grande échelle. Ils doivent donc choisir d'utiliser les sujets ou contenus qui sont mis à leur disposition, et malgré cela l'utilisation de données disponibles présente d'importants attraits. Cependant, ces données comportent une lacune additionnelle pouvant restreindre, de façon importante, la portée des conclusions de l'étude; il s'agit du manque de profondeur de l'information générée. En effet, comme on l'a vu, les questionnaires contextuels se limitent presque exclusivement à sonder les participants au sujet de leur degré d'accord ou de leur fréquence d'utilisation de certaines pratiques pédagogiques. S'il est facile d'envisager que les enseignants n'utilisent pas tous une pratique pédagogique à la même fréquence, ne serait-il pas tout aussi admissible que la façon dont ils conçoivent et appliquent cette pédagogie dans leur salle de classe puisse varier également? Les questionnaires, tel qu'ils sont conçus présentement, ne permettent pas de recueillir de données au sujet de « *comment* » les différentes pratiques pédagogiques sont appliquées par les enseignants. Pensez à vos années passées sur les bancs d'école, est-ce que tous vos enseignants appliquaient la même pratique pédagogique de façon identique ? Étaient-ils tous aussi efficaces dans leur application ? Et, est-ce que cela engendrait exactement le même effet sur l'apprentissage des élèves ? Mes propres expériences me laissent penser le contraire. Cet aspect important vient peut-être expliquer pourquoi, avec les données

actuelles, il est si difficile de lier les pratiques pédagogiques des enseignants avec le rendement des élèves. Il serait logique de croire que la façon dont un enseignant applique une pratique pédagogique donnée pourrait avoir autant, si non plus, d'impact sur le rendement des élèves que la fréquence à laquelle il applique cette pratique. Et donc, pour toutes les raisons mentionnées déjà, il serait faux de croire que les dirigeants peuvent s'informer de façon adéquate au sujet de ce qui se passe dans les salles de classes et de comment cela influence véritablement les apprentissages des élèves en ayant recours exclusivement aux données de questionnaires contextuels d'évaluations à grande échelle.

Limites de l'étude

Certes, la plus grande limite de l'étude fût d'avoir à élaborer une méthodologie de recherche pièce par pièce car aucune étude semblable n'a pu être répertoriée. Plusieurs études ont été consultées et des composantes de leur méthodologie pouvant être appliquée dans le cadre de la présente étude ont été retenues. Nos données impliquaient des contraintes importantes comme le fait de ne pas pouvoir sonder les répondants originaux pour une deuxième prise d'information afin d'établir la fidélité des questionnaires. Les méthodologies utilisées ont dû être choisies, adaptées et parfois développées afin de pouvoir rencontrer les exigences de l'étude. Le résultat final est un mélange éclectique de diverses méthodes qualitatives et quantitatives servant à décrire différents aspects de la qualité psychométrique des items.

Dans ce processus, un instrument de mesure (questionnaire) a été élaboré afin de recueillir des renseignements auprès d'un panel d'enseignants volontaires. Faute de temps et de personnes volontaires, ce questionnaire n'a pas été mis à l'essai avant son utilisation avec

les participants de l'étude. Par conséquent, certaines questions n'ont pas toujours été bien interprétées par les panélistes. De plus, les participants ont préféré répondre au questionnaire individuellement et par écrit plutôt qu'en présence de la chercheuse tel que prévu initialement. Ce premier scénario aurait pu permettre de clarifier l'interprétation de certaines des questions au besoin.

Les échantillons d'enseignants sont de tailles importantes pour la population étudiée, toutefois, ils ne sont pas représentatifs de ces populations. Certaines caractéristiques socio-démographiques ont été examinées afin de déterminer l'adéquation des ces échantillons. Les résultats indiquent que les observations ne sont donc pas généralisables à l'ensemble de la population des enseignants de mathématiques dans les écoles françaises de l'Ontario.

Conclusion

L'objectif principal de la présente étude était d'examiner différents aspects de la qualité psychométrique d'items contextuels portant sur les nouvelles pratiques pédagogiques des enseignants. Ce premier objectif fût ciblé dans le but précis de pouvoir répondre à une préoccupation plus subtile, soit de déterminer si les données générées par les items contextuels des programmes d'évaluation à grande échelle possèdent suffisamment de validité et de fidélité pour servir à informer les décideurs au sujet des changements qui s'opèrent de nos jours, dans les salles de classes.

Les analyses qualitatives et quantitatives réalisées lors de cette étude ont permis de tirer des conclusions intéressantes. D'abord, les résultats des analyses liées à la validité (apparente, de contenu, et théorique, la concordance élève/enseignant et le biais de réponse) laissent présager un manque de validité des données contextuelles relevées par les questionnaires de la TEIMS au sujet des pratiques pédagogiques émergentes. Les lacunes en terme de validité d'apparence et de contenu proviennent principalement d'aspects reliés à la conception des items. Il s'agit en effet de faiblesses au niveau de la formulation de la question, du vocabulaire utilisé, de la période de référence de l'item ou encore de la pertinence du sujet et des choix de réponse pour les répondants. La concordance entre élève et enseignant a permis de confirmer que les items simples et clairement interprétés produisent des résultats plus comparables que les items plus complexes faisant appel à des concepts vagues ou à un trop grand nombre de concepts. La validité théorique n'a pu être établie en fonction des regroupements théoriques effectués correspondant aux principales compétences ciblées par les recommandations du curriculum de l'Ontario de 1997. Les

propriétés métriques s'avèrent également problématiques compte tenu des limites sérieuses imposées par le type de données générées et les lacunes distributives des items. Toutefois, les analyses n'ont pas permis de déceler la présence d'un biais de réponse systématique en fonction du phénomène de désirabilité sociale tant discuté dans les écrits. Les analyses de cohérence interne n'ont pas réussi à démontrer une grande cohésion entre les items des questionnaires qui traitent des nouvelles pratiques et croyances pédagogiques. Ces résultats indiquent en grande partie que les données provenant de questionnaires contextuels semblables à ceux utilisés par les nombreux programmes à grande échelle n'arrivent pas à dresser un portrait valide ni fidèle des nouvelles pratiques pédagogiques chez les enseignants. Les résultats de cette étude doivent cependant, être interprétés avec discernement à cause de certaines faiblesses méthodologiques, dont les données non-représentatives de la population ciblée et le choix limité des moyens méthodologiques pouvant être employés pour démontrer les différents concepts de validité et de fidélité.

Pour conclure, l'acte d'enseigner est incroyablement complexe, il fait intervenir de multiples facteurs et ses fondements théoriques ainsi que son application quotidienne évoluent constamment. C'est sans doute pourquoi il est difficile d'en obtenir une représentation adéquate. Une grande part des connaissances détenues par les gens du domaine de l'éducation provient de recherches réalisées à plus petite échelle, i.e. dans quelques classes seulement. Ces études offrent une plus grande profondeur d'information mais leurs résultats ne sont pas généralisables à l'ensemble d'une population. Ce qui vient donner la raison d'être des évaluations à grande échelle, qui ont malgré leurs limites une utilité particulière, i.e. celle d'informer sur les tendances qui se manifestent globalement au sein d'une population. Les études de plus petite envergure, principalement de nature qualitative, doivent continuer à guider l'élaboration du contenu des questionnaires à grande

échelle. Ces deux types d'études sont nécessaires pour obtenir un portrait valide et complet au sujet des changements qui s'opèrent au niveau des croyances et pratiques pédagogiques des enseignants.

RÉFÉRENCES

- Anderson, J., Rogers, T., Ungerleider, C. (juin, 2004). *Preliminary Models of School Performance*. Communication personnelle, 2004.
- Bégin, J. (2006). Introduction à la psychométrie, Département de psychologie Université du Québec à Montréal. Information obtenue le 1^{er} aout 2006 à partir du site : <http://www.er.uqam.ca/nobel/r30574/>
- Blank, R., et Pechman, E. (1995). *State Curriculum Frameworks in Mathematics and Science : How Are They Changing Across the States ?* Washington, DC: Council of Chief State School Officers.
- Burstein, L. (1995). *Validating National Curriculum Indicators*, RAND Corporation and the National Science Foundation, Santa-Monica, Californie.
- Cheng , S. K. et Seng , Q.K. (2001). Sex differences in TIMSS mathematics achievement of four asian nations. *Studies in Educational Evaluation* , 27(4) , 331-340.
- Clare, L. (2000). *Using teachers' assignments as an indicator of classroom practice*. Center for the Study of Evaluation Technical Report #532. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Clauser, P. S. (2001). *Effects of educational policy: School characteristics and student background on Hispanic and Anglo students' NAEP reading proficiency*. Unpublished doctoral thesis ED3033936. Albuquerque: University of New Mexico.
- Crocker, R. K. (2002). *Résultats d'apprentissage : Analyse critique du domaine au Canada*. Conseil des statistiques canadiennes de l'éducation.

- Darling-Hammond, L. & Falk, B. (1997). Using standards and assessments to support student learning. *Phi Delta Kappan*, 79, 190-199.
- Davidson, A. B. (1999). Eighth-grade science teachers use of instructional time: Examining questions from the Third International Mathematics and Science Study (TIMSS) and comparing TIMSS and National Science Foundation questionnaires. *Digital Dissertations*, (Numéro de publication : AAT 9933322). Texte obtenu du site Internet de Digital Dissertations.
- Dechang, C., Zhenqiu, L., Xiaobin, M., & Dong, H. (2005). Selecting genes by test statistics. *Journal of Biomedicine and Biotechnology*, 2005(2), 132–138. Consulté le 24 octobre 2006, à partir du site : <http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1184045&blobtype=pdf>
- Fowler, F. & Poetter, T. (2004) Framing French Success in Elementary Mathematics: Policy, Curriculum, and Pedagogy, *Curriculum Inquiry*, 34(3), 283-314.
- Foy, P. & Joncas, M. (2001). PIRLS Sampling Design. In Martin, M.O., Mullis, I.V.S., & Kennedy, A.M. (Eds.), *PIRLS 2001 Technical Report* (pp. 53-66). Chestnut Hill, MA: Boston College.
- Fraenkel, J. R. & Wallen, N.E. (2003). *How to design and evaluate research in education*, 5e édition, New York : McGraw-Hill.
- Fuhrman, S. (2003). Riding waves, trading horses. The twenty-year effort to reform education. In D.T. Gordon (Ed.), *A nation reformed? American education 20 years after A Nation At Risk* (pp. 7-22). Cambridge, MA: Harvard Education Press.
- Herry, Y., (2000). Apprentissage et enseignement des sciences : Résultats de la troisième enquête internationale. *Revue des sciences de l'éducation*, 26, 347-366.

- Gales, M. J. (2000). Relationship between constructivist teacher beliefs and instructional practices to students' mathematical achievement: Evidence from TIMSS. *Digital Dissertations*, (Numéro de publication : AAT 9959146). Texte obtenu du site Internet de Digital Dissertations.
- Giguère, D. (1999, Juin). Gender Gap Widening Among Ontario Teachers. *Professionally Speaking*, consulté le 24 avril 2006, http://www.oct.ca/publications/professionally_speaking/june_1999/gap.htm
- Gorsuch, R. L. (1997). Exploratory Factor Analysis : Its Role in Item Analysis. *Journal of personality assessment*, 68(3), 532-560.
- Grouws, D. A. & Cebulla, K. J. (2000). *Improving student achievement in mathematics*. Lausanne, Switzerland : International Academy of Education(IAE) et le Bureau International de l'Éducation-UNESCO.
- Hill, C. L. (2002). What are the characteristics of effective teaching? A comparative study of stakeholder perceptions. *Digital Dissertations*, (Numéro de publication : AAT 3053200). Texte obtenu du site Internet de Digital Dissertations.
- Hodges, E. L. (2000). Influence of number of topics, topic duration, and curricular focus on biology achievement of Population 3 TIMSS countries. *Digital Dissertations*, (Numéro de publication : AAT 9998483). Texte obtenu du site Internet de Digital Dissertations.
- Howell, D.C. (1998). *Méthodes statistiques en sciences humaines*, Paris : DeBoeck Université.
- Hu, L. & Bentler, P., M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 4, 1-55.

- Johnson, B. & Christensen, L. (2004). *Educational research : quantitative, qualitative and mixed approaches*, 2^e édition, Toronto: Allyn & Bacon.
- Landry, R. & Allard, R. (2002, Octobre). *Résultats pancanadiens des élèves francophones en milieu minoritaires aux évaluations du PIRS: variables contextuelles et conséquences pédagogiques* (University of Moncton). Obtenu le 15 novembre 2005 , à partir du site du Conseil des Ministres de l'éducation du Canada : [http://www.cmec.ca/else/francophone/Rapport TechniqueVoletA.fr.pdf](http://www.cmec.ca/else/francophone/Rapport%20TechniqueVoletA.fr.pdf)
- Lawson, A. C. (2002). Teachers' reported understanding and implementation of a new grade 8 mathematics curriculum. *Digital Dissertations*, (Numéro de publication : AAT NQ75589). Texte obtenu du site Internet de Digital Dissertations.
- Legendre, R. (1993). *Dictionnaire actuel de l'éducation*, 2^e édition. Montréal : Guérin Éditeurs limitée.
- Legendre, R. (2002). *Stop aux réformes scolaires*. Montréal : Guérin Éditeurs limitée.
- Linn, R. L., & Baker, E. L. (1998, Automne). School quality: Some missing pieces. *CRESST Line*. Los Angeles : University of California, National Center for research on Evaluation, Standards, and Student Testing (CRESST).
- Loewenberg-Ball, D. L., & Rowan, B. (2004). Introduction: Measuring instruction. *Elementary School Journal*, 105 (1), 3-10.
- Martin, M.O., Mullis, I.V.S., & Chrostowski, S.J. (2004). *TIMSS 2003 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Obtenu le 12 avril 2005 à l'adresse suivante : <http://timss.bc.edu/timss2003i/technicalD.html>
- Masini, B. & Taylor, J. (Décembre, 2000). New evidence links curricula and instruction to mathematics achievement. *Policy Issues*, 7, 1-6.

Mayer, D.P. (1999). Measuring Instructional Practice : Can policy makers trust survey data ? *Educational Evaluation and Policy Analysis*, 21 (1), 29-45.

McIntyre, F. (1998, Décembre). Pénurie d'enseignants à l'horizon. *Pour parler profession*, Consulté le 25 avril 2006 à l'adresse suivante:

http://www.oct.ca/publications/pour_parler_profession/decembre_1998/cover.htm

Ministère de l'Éducation et de la Formation de l'Ontario. (1997). Le curriculum de l'Ontario de la 1^{re} à la 8^e année : Mathématiques. Obtenu le 27 octobre à l'adresse suivante :
<http://www.edu.gov.on.ca/fre/document/curricul/curr97mf.pdf>

Government of Alberta. (1997). *K-12 Learning System Policy, Regulations and Forms Manual*, Directive 4.2.1 - Teaching Quality Standard Applicable to the Provision of Basic Education in Alberta. Texte obtenu le 2 août 2006 à l'adresse suivante :
<http://www.education.gov.ab.ca/educationguide/pol-plan/polregs/adobe/421.pdf>

Mullens, J., & Kasprzyk, D. (1999). *Validating item response on self-report teacher surveys: NCES working paper*. Washington, DC: U.S. Department of Education. Texte obtenu le 5 janvier 2005 à l'adresse suivante :
http://www.amstat.org/sections/srms/proceedings/papers/1999_118.pdf

Mullis, I., (2002). Background Questions in TIMSS and PIRLS: An Overview, Paper Commissioned by the National Assessment Governing Board. Obtenu le 11 novembre 2004, à l'adresse suivante : <http://www.nagb.org/release/Mullis.doc>

Muthen, L.K.; Muthen, B.O. (année...1998) *Mplus User's Guide*, Third Edition. Los Angeles, CA: Muthen & Muthen, 1998-2005.

National Center for Education Statistics (NCES), U.S. Department of Education. (1999). *Measuring Classroom Instructional Processes: Using Survey and Case Study*

- Fieldtest Results to Improve Item Construction. Working Paper No. 1999-08.
- Mullens, J., Gayler, K., et Kasprzyk, D. Washington, D.C.
- National Center for Education Statistics (NCES), U.S. Department of Education. (2002).
The Measurement of Instructional Background Indicators: Cognitive Laboratory
Investigations of the Responses of Fourth and Eighth Grade Students and Teachers to
Questionnaire Items. Working Paper No. 1999-08. Levine, R., Huberman, M.,
Buckner, K., Goldstein, A. Washington, D.C.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards
for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- National Council of Teachers of Mathematics. (1991). *Professional standards for teaching
mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- National Council of Teachers of Mathematics (2000). *Principals of Standards for School
Mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- O'Connor-Petruso, S., Shiering, M., Hayes, B. et Serrano, B. (Mai, 2004). *Pedagogical and
Parental Influences in Mathematics Achievement by Gender Among Select European
Countries From the TIMSS-R Study*. Document présenté lors de la 1re Conférence
Internationale de recherche de l'IEA. Lefkosia, Chypre.
- Organisation de coopération et de développement économique (OCDE). (1995). *La qualité
de l'enseignement*. Paris : Les Éditions de l'OCDE.
- Pelgrum, W.J., Voogt, J., & Plomp, T. (1995). Curriculum indicators in international
comparative research. Dans : *Measuring the quality of schools*. Paris : OECD.
- Pepermans, R. (1990). *Vocabulaire de l'administration publique et de la gestion*. Ottawa :
Secrétariat d'État du Canada.

- Porter, A., & Smithson, J. (2001). Defining, developing, and using curriculum indicators. Research Report Series RR-048. Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania.
- Puk, T. (1999). Formula for success according to the TIMSS or the subliminal decay of jurisdictional educultural integrity? *Alberta Journal of Educational Research*. 45(3), 222- 232.
- Rey B., Coche, F., Kahn, S., Robin, F. (août 2005). *Pratique pédagogique à l'école primaire et réussite scolaire des élèves venant de milieux défavorisés*. Rapport de recherche, Communauté française, Service des Sciences de l'Éducation, Bruxelles.
- Robitaille, D.F., & Garden, R.A. (1996). *Research Questions & Study Design*. TIMSS Monograph No.2 Vancouver: Pacific Educational Press.
- Robitaille, D. F., & Taylor, A. R. (2002). *Rapport du Canada TÉIMS-99: Sommaire*. Tiré le 15 septembre 2003 du site <http://www.curricstudies.educ.ubc.ca/wprojects/TIMSS/Summary-FRE.pdf>
- Rodriguez, M. C. (1999). Linking classroom assessment practices to large-scale test performance. *Digital Dissertations*, (Numéro de publication : AAT 9948164). Texte obtenu du site Internet de Digital Dissertations.
- Rogers, T. W., Anderson, J., Klinger, D. A., & Dawber, T. (2004). Pitfalls and potential of secondary data analysis of the council of ministers of education, Canada National Assessments, *Canadian Educational Research Association / Canadian Society for Studies in Education*.Winnipeg.
- Rowan, B., Correnti, R., and Miller, R. (2002). What Large-Scale Survey Research Tells Us About Teacher Effects on Student Achievement: Insights from the Prospects Study of Elementary Schools. *Teachers College Record*, 104(8), 1525-1567.

- Scheerens, J. (2000). *Améliorer l'efficacité des écoles*. Paris: UNESCO, IPE, Principes de la planification de l'éducation, numéro de série 68.
- Schmidt, W. H. and Cogan, L.S. (1996). Development of the TIMSS Context Questionnaires in M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study (TIMSS) Technical Report, Volume I: Design and Development*. Chestnut Hill, MA: Boston College.
- Simon, M., Tierney, R., Plouffe, S., (Mai, 2005). *Hindrances and Solutions in Minority Group Research Using International Assessment Data*. Communication présentée dans le cadre de la conférence annuelle de la Société Canadienne pour l'Étude de l'Éducation (SCÉÉ), London, Ontario.
- Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis*, 21(1), 1-27.
- SPSS for Windows, Version 11.0.0. (2001). En version standard, Chicago: SPSS Inc.
- StatSoft, Inc. (2006). Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB:
<http://www.statsoft.com/textbook/stathome.html>.
- Stecher, B., Hamilton, L., Ryan, G., Le, V.-N., Williams, V., Robyn, A., et al. (avril, 2002). *Measuring reform-oriented instructional practices in mathematics and science*. Communication présentée dans le cadre de la conférence annuelle du American Educational Research Association, New Orleans. LA.
- Stigler, J., & Hiebert, J. (1997). Understanding and improving classroom mathematics instruction: An overview of the TIMSS Video Study. *Phi Delta Kappan*, 79(1), 14-21.

- Tardif, J. (1992). *Pour un enseignement stratégique: l'apport de la psychologie cognitive*.
Montréal : Éditions Logiques.
- Tharp, R.G., & Gallimore, R. (1988). *Rousing Minds to Life: Teaching, Learning, and
Schooling in Social Context*. Cambridge: Cambridge University Press.
- Ungerleider, C. (2003). Large-Scale Student Assessment : Guidelines for Policymakers.
International Journal of testing, 3(2), 119-128.
- Webster , B. J., et Fisher , D. L. (2000). Accounting for variation in science and
mathematics achievement: A multilevel analysis of Australian data; Third
International Mathematics and Science Study (TIMSS). *School Effectiveness and
School Improvement* , 11(3) , 339-360.

ANNEXES

Annexe A

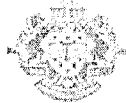
**Inventaire des items sélectionnés à partir des
questionnaires contextuels de la TEIMS (1995, 1999 et 2003)**

N° de l'item	Thème abordé par la question	Items Enseignant			Item Élève		
		1995	1999	2003	1995	1999	2003
1	Importance de mémoriser des formules et des méthodes	X	X	X	X	X	
2	Utiliser plusieurs mode de représentation	X	X	X			
3	Un ensemble d'algorithmes ou de règles	X	X	X			
4	Expliquer le raisonnement qui sous-tend une idée	X	X	X			X
5	Utiliser la calculatrice pendant les cours de math	X	X	X	X	X	X
6	Présenter et analyser, tableaux, diagrammes, graphiques	X	X	X			X
7	Utiliser la calculatrice pour des problèmes complexes	X	X	X			
8	Trouver des équations pour représenter des relations	X	X	X			X
9	Travailler par deux ou en petits groupes	X	X	X	X	X	X
10	Résoudre des problèmes sans solution évidente	X	X	X			
11	Utiliser un ordinateur pour problèmes ou exercices	X	X	X	X	X	
12	Comprendre l'application des math dans la vie courante	X	X	X	X	X	X
13	Les math constituent un guide pour les situations réelles	X	X	X			
14	Calculatrice pour explorer les notions numériques	X	X	X			
15	Le manque de matériel informatique	X	X	X			
16	Élève doit découvrir des applications du sujet étudié	X	X	X			
17	Élève doit expliquer son raisonnement lors de tests	X	X	X			
18	Penser de façon ordonnée et séquentielle pour les math	X	X				
19	Comprendre les notions et les méthodes des math	X	X				
20	Penser de façon créatrice pour les math	X	X				
21	Les math = façon structurée de représenter le monde réel	X	X				

22	Petites recherches ou collecte de données comme devoir	X	X				
23	Fournir des raisons à l'appui de leurs solutions	X	X				
24	Résolution faite par essais qui modifient nos conclusions			X			
25	Différentes façons de résoudre les problèmes de math			X			
26	Temps consacré à résoudre des problèmes seul			X			X
27	Déterminer sa façon de résoudre un problème complexe			X			X
28	Élèves ont accès aux ordinateurs pour les cours de math			X			
29	Utiliser ordinateur pour démontrer des concepts math			X			
30	Utiliser ordinateur pour chercher des idées et de l'info			X			X
31	Utiliser ordinateur pour traiter et analyser des données			X			X
32	Questions de test sur les régularités et les relations			X			
TOTAL		23	23	26	5	5	10

Annexe B

**Certificat d'approbation déontologique émis par le Comité d'éthique de la recherche
en Sciences sociales et humanités de l'Université d'Ottawa – 20 janvier 2006**



Université d'Ottawa University of Ottawa

Servicio de subvenciones de investigación y deontología Research Grants and Ethics Services

**COMITÉ D'ÉTHIQUE DE LA RECHERCHE
EN SCIENCES SOCIALES ET HUMANITÉS**

ATTESTATION D'APPROBATION DÉONTOLOGIQUE

La présente attestation certifie que le Comité d'éthique de la recherche en Sciences Sociales et Humanités de l'Université d'Ottawa a examiné la demande d'approbation déontologique pour le projet intitulé **Analyse d'Items du questionnaire contextuel des enseignants de la TEIMS portant sur les pratiques pédagogiques (Dossier # 08-05-08)**, présenté par Sarah Plouffe et supervisé par Marielle Simon de la Faculté d'éducation. Le Comité d'éthique a déterminé que la demande respectait les principes déontologiques établis par l'Énoncé de politique des trois conseils et par les règles de procédure des Comités d'éthique de l'Université d'Ottawa et a donc accordé une catégorie 1a (approbation) à ce projet.

La présente attestation est valide pour un an à partir de la date indiquée ci-dessous.

Catherine Paquet
Responsable de l'éthique en recherche
Pour le Président du CÉR en Sciences Sociales et Humanités
Richard Clément

20 janvier 2006

Date

Annexe C

**Tableau des items sélectionnés à partir des questionnaires enseignants
de la TEIMS 1995 – 1999 - 2003**

Items TEIMS retenus pour les analyses

1995		1999		2003	
Questionnaires des enseignants	Questionnaire des élèves	Questionnaires des enseignants	Questionnaire des élèves	Questionnaires des enseignants	Questionnaire des élèves
BTBMIMP1 Q. 15-A Pour qu'ils réussissent en mathématiques à l'école, dans quelle mesure pensez-vous qu'il est important pour les élèves... de mémoriser des formules et des méthodes? CHOIX : pas important, plus ou moins important, très important.	BSBMDOW4 Q. 19-D Pour bien réussir en mathématiques, il faut... Apprendre par cœur le manuel scolaire ou les notes CHOIX : Entièrement d'accord, d'accord, pas d'accord, Pas du tout d'accord	BTBMIMP1 Q. 11-A Pour qu'ils réussissent en mathématiques à l'école, dans quelle mesure pensez-vous qu'il est important pour les élèves... De mémoriser des formules et des méthodes. CHOIX : pas important, plus ou moins important, très important.	BSBMDOW4 Q. 18-D Pour bien réussir en mathématiques, il faut... Apprendre par cœur le manuel scolaire ou les notes CHOIX : Entièrement d'accord, d'accord, pas d'accord, Pas du tout d'accord	BTBMADME Q. 14-D Dans quelle mesure êtes-vous d'accord ou en désaccord avec les énoncés suivants? Apprendre les mathématiques est surtout une question de mémoire. CHOIX : Tout à fait d'accord, d'accord, en désaccord, tout à fait en désaccord.	
Commentaires : Question et choix de réponse sont identiques pour 1995 et 1999. Mais la formulation de la question et les choix de réponse sont légèrement différents pour 2003. Le nombre de catégories de réponse passe de 3 à 4.					
BTBMAGR6 Q.16-F Dans quelle mesure êtes-vous d'accord avec chacun des énoncés suivants ? Pour enseigner un sujet en mathématiques, il faut faire appel à plus d'un mode de représentation (illustration, matériel concret, ensemble de symboles, etc.) CHOIX : Fortement en désaccord, en désaccord, d'accord, entièrement d'accord.		BTBMAGR6 Q.12-F Dans quelle mesure êtes-vous d'accord avec chacun des énoncés suivants ? Pour enseigner un sujet en mathématiques, il faut faire appel à plus d'un mode de représentation (illustration, matériel concret, ensemble de symboles, etc.) CHOIX : Fortement en désaccord, en désaccord, d'accord, entièrement d'accord.		BTBMADMR Q. 14-A Dans quelle mesure êtes-vous d'accord ou en désaccord avec les énoncés suivants? Il faut plus d'une représentation (dessin, matériel concret, symbole, ensemble, etc.) pour enseigner les mathématiques. CHOIX : Tout à fait d'accord, d'accord, en désaccord, tout à fait en désaccord.	
Commentaires : Les choix de réponse sont un peu différents entre (1995 et 1999) et 2003, l'échelle est inversée. La formulation de la question est un peu différente pour 2003.					

BTBMAGR7		BTBMAGR7		BTBMADSA	
Q.16-G	Dans quelle mesure êtes-vous d'accord avec chacun des énoncés suivants ? Les mathématiques doivent être apprises comme un ensemble d'algorithmes ou de règles qui couvrent toutes les possibilités. CHOIX : Fortement en désaccord, en désaccord, d'accord, entièrement d'accord.	Q.12-G	Dans quelle mesure êtes-vous d'accord avec chacun des énoncés suivants ? Les mathématiques doivent être apprises comme un ensemble d'algorithmes ou de règles qui couvrent toutes les possibilités. CHOIX : Fortement en désaccord, en désaccord, d'accord, entièrement d'accord.	Q.14-B	Dans quelle mesure êtes-vous d'accord ou en désaccord avec les énoncés suivants? Les mathématiques devraient s'apprendre comme des ensembles d'algorithmes ou de règles couvrant toutes les possibilités. CHOIX : Tout à fait d'accord, d'accord, en désaccord, tout à fait en désaccord.
Commentaire : Choix de réponse légèrement différents. La question est presque identique pour les 3 cycles de la TEIMS.					

BTBMASK1		BTBMASK1		BTBMASEA		BSBMHEXP	
Section B Q. 15-A	Dans vos cours de mathématiques, à quelle fréquence demandez-vous aux élèves de réaliser chacune des activités suivantes ? Expliquer le raisonnement qui sous-tend une idée. CHOIX : Jamais ou presque jamais, quelquefois, la plupart du temps, toutes les fois.	Section B Q.10-A	Pendant les cours de mathématiques, vous arrive-t-il de demander aux élèves de réaliser les activités suivantes ? Expliquer le raisonnement qui sous-tend une idée. CHOIX : Jamais ou presque jamais, parfois, la plupart du temps, à chaque fois.	Q. 21-H	À quelle fréquence demandez-vous aux élèves de mathématiques de la classe TEIMS de faire ce qui suit? Expliquer le raisonnement qui sous-tend une idée. CHOIX : Tous les cours, la moitié des cours, quelques cours, jamais.	Q. 10-G	À quelle fréquence faites-vous les activités ci-dessous pendant les cours de mathématiques ? Nous expliquons nos réponses CHOIX : Presque chaque cours, Environ la moitié des cours, certains cours, jamais
Commentaires : Formulation de la question est semblable entre (1995 et 1999) et 2003. Les choix de réponse sont légèrement différents entre (1995 et 1999) et 2003.							

BTBMCALC		BSBMCALC		BTBMCAML		BSBMHCAL	
Section B Q. 8	Pendant les cours de mathématiques, combien de vos élèves ont accès à une calculatrice ?	Q. 25-F	Ces événements se produisent-ils souvent pendant les cours de mathématiques ? Nous expliquons nos réponses	Section B Q. 25	Dans quelle mesure les élèves de la classe TEIMS sont-ils autorisés à utiliser une calculatrice pendant les cours de mathématiques?	Q. 10-N	À quelle fréquence faites-vous les activités ci-dessous pendant les cours de mathématiques ? Nous utilisons des calculatrices
Section B Q. 26-F Les situations suivantes arrivent-elles fréquemment pendant les cours de mathématiques ? Nous utilisons des calculatrices.							

BTBMSK5	BTBMSK5	BTBMSRR	BSBMHEFR
Section B Q. 15-E Dans vos cours de mathématiques, à quelle fréquence demandez-vous aux élèves de réaliser chacune des activités suivantes ? Indiquer des équations pour représenter des rapports. CHOIX : Jamais ou presque jamais, quelquefois, la plupart du temps, toutes les fois.	Section B Q. 10-E Pendant les cours de mathématiques, vous arrive-t-il de demander aux élèves de réaliser les activités suivantes ? Trouver des équations pour représenter des relations. CHOIX : Jamais ou presque jamais, parfois, la plupart du temps, à chaque fois.	Q. 21-E À quelle fréquence demandez-vous aux élèves de mathématiques de la classe TEIMS de faire ce qui suit? Écrire des équations et des fonctions qui représentent des relations CHOIX : Tous les cours, la moitié des cours, quelques cours, jamais.	Q. 10-D À quelle fréquence faites-vous les activités ci-dessous pendant les cours de mathématiques ? Nous écrivons des équations et des fonctions pour représenter des relations CHOIX : Presque chaque cours, Environ la moitié des cours, certains cours, jamais
Commentaires : La formulation de la question change un peu d'un cycle à l'autre. Les choix de réponse sont différents entre (1995-1999) et 2003.			

BTBMSK6	BSBMSGRP	BTBMS6	BSBMSGRP	BTBMASSG	BSBMHWSG
Q. 17-F Dans les cours de mathématiques, à quelle fréquence les élèves doivent-ils... travailler deux par deux ou en petits groupes, avec l'aide de l'enseignant(e). CHOIX : jamais ou presque jamais, quelquefois, la plupart du temps, toutes les fois.	Q. 25-H Ces événements se produisent-ils souvent pendant les cours de mathématiques ? Nous travaillons par paires ou en petits groupes CHOIX : Presque toujours, Assez souvent, parfois, jamais.	Section B Q. 11-F Pendant les cours de mathématiques, arrive-t-il aux élèves de ... Travailler par deux ou en petits groupes, avec l'aide de l'enseignant ? CHOIX : jamais ou presque jamais, quelquefois, la plupart du temps, toutes les fois.	Section B Q. 26-i Les situations suivantes arrivent-elles fréquemment pendant les cours de mathématiques ? Nous travaillons à deux ou en petits groupes CHOIX : Presque toujours, Assez souvent, à l'occasion, jamais.	Q. 21-F À quelle fréquence demandez-vous aux élèves de mathématiques de la classe TEIMS de faire ce qui suit? Travailler en petits groupes CHOIX : Tous les cours, la moitié des cours, quelques cours, jamais.	Q. 10-E À quelle fréquence faites-vous les activités ci-dessous pendant les cours de mathématiques ? Nous travaillons en petits groupes CHOIX : Presque chaque cours, Environ la moitié des cours, certains cours, jamais
Commentaires :					

BTBMSK3	BTBMSK3	BTBMASWS
Section B Q. 15-C Dans vos cours de mathématiques, à quelle fréquence demandez-vous aux élèves de réaliser chacune des activités suivantes ? Travailler sur des problèmes pour lesquels il n'y a aucune solution immédiate évidente.	Section B Q. 10-C Pendant les cours de mathématiques, vous arrive-t-il de demander aux élèves de réaliser les activités suivantes ? Travailler sur des problèmes pour lesquels il n'y a aucune solution immédiate évidente.	Q. 21-C À quelle fréquence demandez-vous aux élèves de mathématiques de la classe TEIMS de faire ce qui suit? Résoudre des problèmes dont la solution n'est pas immédiatement perceptible

CHOIX : Jamais ou presque jamais, quelquefois, la plupart du temps, toutes les fois.	CHOIX : Jamais ou presque jamais, parfois, la plupart du temps, à chaque fois.	CHOIX : Tous les cours, la moitié des cours, quelques cours, jamais.
Commentaires :		

BTBMASK4		BSBMCOMP		BTBMASK4		BSBMCOMP		BTBMCASP	
Section B Q. 15-D Dans vos cours de mathématiques, à quelle fréquence demandez-vous aux élèves de réaliser chacune des activités suivantes ? Se servir d'un ordinateur pour résoudre des problèmes ou faire des exercices.	Section B Q. 15-D Ces événements se produisent-ils souvent pendant les cours de mathématiques ? Nous utilisons des ordinateurs	Section B Q. 10-D Pendant les cours de mathématiques, vous arrive-t-il de demander aux élèves de réaliser les activités suivantes ? Utiliser un ordinateur pour résoudre des problèmes ou faire des exercices.	Section B Q. 26-G Les situations suivantes arrivent-elles fréquemment pendant les cours de mathématiques ? Nous utilisons des ordinateurs	Q. 31-B À quelle fréquence utilisez-vous l'ordinateur pour les activités ci-dessous, dans vos cours de mathématiques de la classe TEIMS? Permettre aux élèves de mettre habiletés et méthodes en pratique	CHOIX : Jamais ou presque jamais, quelquefois, la plupart du temps, toutes les fois.	CHOIX : Jamais ou presque jamais, parfois, la plupart du temps, à chaque fois.	CHOIX : Presque toujours, Assez souvent, parfois, jamais.	CHOIX : Presque toutes les cours, La moitié des cours, Quelques cours, Jamais.	
Commentaires : Les questions ne sont pas identiques mais rejoignent le même sujet.									

BTBMIMP5		BSBMLIFE		BTBMIMP5		BSBMLIFE		BTBMASDL		BSBMHMDL	
Q. 15-E Pour qu'ils réussissent en mathématiques à l'école, dans quelle mesure pensez-vous qu'il est important pour les élèves... De comprendre l'application des mathématiques dans la vie courante	Q. 23-D Que penses-tu des mathématiques ? les mathématiques sont importantes dans la vie de tout le monde.	Q. 11-E Pour qu'ils réussissent en mathématiques à l'école, dans quelle mesure pensez-vous qu'il est important pour les élèves... De comprendre l'application des mathématiques dans la vie courante.	Q. 24-D Que penses-tu des mathématiques ? les mathématiques sont importantes dans la vie de tout le monde.	Q. 21-G À quelle fréquence demandez-vous aux élèves de mathématiques de la classe TEIMS de faire ce qui suit? Établir un lien entre ce qu'ils apprennent en mathématiques et leur vie quotidienne.	Q. 10-F À quelle fréquence faites-vous les activités ci-dessous pendant les cours de mathématiques ? Nous établissons des liens entre ce que nous apprenons en mathématiques et la vie quotidienne	CHOIX : pas important, plus ou moins important, très important.	CHOIX : Entièrement d'accord, d'accord, pas d'accord, pas du tout d'accord	CHOIX : Entièrement d'accord, d'accord, pas d'accord, pas du tout d'accord	CHOIX : Tous les cours, la moitié des cours, quelques cours, jamais.	CHOIX : Presque chaque cours, Environ la moitié des cours, certains cours, jamais	
Commentaires : Les questions sont identiques pour 1995 et 1999. La question 2003 est différente puisqu'elle traite de ce que l'enseignant demande à ses élèves. La question 1995 et 1999 demande plutôt ce qui est important pour assurer la réussite des élèves selon les enseignants.											

BTBMAGR3		BTBMAGR3		BTBMADRW	
Q. 16-C	Dans quelle mesure êtes-vous d'accord avec chacun des énoncés suivants ? Les mathématiques constituent un guide pratique et structuré pour aborder les situations réelles. CHOIX : Fortement en désaccord, en désaccord, d'accord, entièrement d'accord.	Q. 12-C	Dans quelle mesure êtes-vous d'accord avec chacun des énoncés suivants ? Les mathématiques constituent un guide pratique et structuré pour aborder les situations réelles. CHOIX : Fortement en désaccord, en désaccord, d'accord, entièrement d'accord.	Q. 14-G	Dans quelle mesure êtes-vous d'accord ou en désaccord avec les énoncés suivants ? Pour enseigner les mathématiques, il est essentiel de modéliser les problèmes du monde réel CHOIX : Tout à fait d'accord, d'accord, en désaccord, tout à fait en désaccord.
Commentaires : Match identique pour 1995 et 1999 mais une question semblable pour 2003.					

BTBMCAL5		BTBMCAL5		BTBMCAL5	
Q. 9-E	À quelle fréquence les élèves de votre classe de mathématiques se servent-ils d'une calculatrice pour les activités suivantes : Explorer les notions numériques CHOIX : Presque tous les jours, 1-2 fois par semaine, 1-2 fois par mois, jamais ou pratiquement jamais.	Section B Q. 8-E	À quelle fréquence les élèves de votre classe se servent-ils d'une calculatrice pour les activités suivantes ? Explorer les notions numériques CHOIX : Presque tous les jours, 1-2 fois par semaine, 1-2 fois par mois, jamais ou pratiquement jamais.	Q. 28-D	À quelle fréquence les élèves de la classe TEIMS utilisent-ils la calculatrice pendant les cours de mathématiques pour les activités suivantes ? Explorer des concepts numériques CHOIX : Tous ou presque tous les cours, La moitié des cours, Quelques cours, Jamais
Commentaires : Bon match entre les cycles. Les choix de réponse varient entre (1995-1999) et 2003.					

BTBMLM08		BTBMLM08		BTBGLT07	
Section B Q. 7-H	Selon vous, dans quelle mesure les éléments suivants restreignent-ils votre façon d'enseigner les mathématiques à votre classe ? Manque de matériel informatique	Section B Q. 5-H	Selon vous, dans quelle mesure les éléments suivants restreignent-ils votre façon d'enseigner les mathématiques ? Manque de matériel informatique	Q. 22-G	À votre avis, dans quelle mesure les éléments ci-dessous limitent-ils votre façon d'enseigner à la classe TEIMS ? Le manque de matériel informatique.

CHOIX : Pas du tout, un peu, beaucoup, énormément.	CHOIX : Pas du tout, un peu, beaucoup, énormément.	CHOIX : Beaucoup, Assez, un peu, pas du tout, ne s'applique pas
Commentaires : La question est presque identique entre les cycles.		

BTBMFINN

Section B Q. 20-H Si vous donnez des devoirs de mathématiques, à quelle fréquence attribuez-vous aux élèves chacun des types de travaux suivants ? détermination d'une ou plusieurs applications pour la matière étudiée. CHOIX : Jamais, rarement, parfois, toujours	Section B Q. 17-H Si vous donnez des devoirs mathématiques à vos élèves, vous arrive-t-il de leur attribuer les types de travaux suivants ? La découverte d'une ou plusieurs applications du sujet étudié. CHOIX : jamais, rarement, parfois, toujours	Section B Q. 35-C À quelle fréquence donnez-vous les devoirs ci-dessous aux élèves de la classe TEIMS ? Trouver une ou plusieurs applications des notions abordées. CHOIX : Presque jamais, parfois, presque toujours.
--	--	--

Commentaires : La question est très semblable entre les cycles. Les choix de réponse passent de 4 à 3.

BTBMWGT2

Section B Q. 22-B Lors de l'évaluation des travaux des élèves de votre classe de mathématiques, quelle importance accordez-vous à chacun des types d'évaluation suivants ? Questionnaires exigeant de brèves réponses réalisés par l'enseignant(e) dans lesquels les élèves doivent décrire ou expliquer leur raisonnement. CHOIX : Pas du tout, un peu, beaucoup, énormément	Section B Q. 19-B Lors de l'évaluation des travaux des élèves de votre classe de mathématiques, quelle importance accordez-vous à chacun des types d'évaluation suivants ? Les questionnaires exigeant des réponses brèves mis au point par l'enseignant(e) dans lesquels les élèves doivent décrire ou expliquer leur raisonnement. CHOIX :Aucune, un peu, beaucoup, énormément.	Section B Q. 39-C À quelle fréquence posez-vous les types de questions ci-dessous dans vos tests ou épreuves de mathématiques ? Questions exigeant des explications ou des justifications. CHOIX : Presque jamais, parfois, presque toujours.
--	---	--

Commentaires : Bon match... pour les cycles.

BTBMIMP2	
<p>BTBMIMP2</p> <p>Q. 15-B</p> <p>Pour qu'ils réussissent en mathématiques à l'école, dans quelle mesure pensez-vous qu'il est important pour les élèves... De penser de façon ordonnée et séquentielle.</p> <p>CHOIX : pas important, plus ou moins important, très important.</p>	<p>BTBMIMP2</p> <p>Q. 11-B</p> <p>Pour qu'ils réussissent en mathématiques à l'école, dans quelle mesure pensez-vous qu'il est important pour les élèves... De penser de façon ordonnée et séquentielle.</p> <p>CHOIX : pas important, plus ou moins important, très important.</p>
<p>Commentaires : Match identique pour 1995 et 1999 mais aucun item correspondant pour 2003</p>	

BTBMIMP3	
<p>BTBMIMP3</p> <p>Q. 15-C</p> <p>Pour qu'ils réussissent en mathématiques à l'école, dans quelle mesure pensez-vous qu'il est important pour les élèves... De comprendre les notions, les principes et les méthodes des mathématiques.</p> <p>CHOIX : pas important, plus ou moins important, très important.</p>	<p>BTBMIMP3</p> <p>Q. 11-C</p> <p>Pour qu'ils réussissent en mathématiques à l'école, dans quelle mesure pensez-vous qu'il est important pour les élèves... De comprendre les notions, les principes et les méthodes des mathématiques.</p> <p>CHOIX : pas important, plus ou moins important, très important.</p>
<p>Commentaires : Match identique pour 1995 et 1999 mais aucun item correspondant pour 2003</p>	

BTBMIMP4	
<p>BTBMIMP4</p> <p>Q. 15-D</p> <p>Pour qu'ils réussissent en mathématiques à l'école, dans quelle mesure pensez-vous qu'il est important pour les élèves... De penser de façon créatrice.</p> <p>CHOIX : pas important, plus ou moins important, très important.</p>	<p>BTBMIMP4</p> <p>Q. 11-D</p> <p>Pour qu'ils réussissent en mathématiques à l'école, dans quelle mesure pensez-vous qu'il est important pour les élèves... De penser de façon créatrice.</p> <p>CHOIX : pas important, plus ou moins important, très important.</p>
<p>Commentaires : Match identique pour 1995 et 1999 mais aucun item correspondant pour 2003</p>	

BTBMAGR2

BTBMAGR2

<p>Q. 16-B Dans quelle mesure êtes-vous d'accord avec chacun des énoncés suivants ? Les mathématiques sont principalement une façon structurée de représenter le monde réel. CHOIX : Fortement en désaccord, en désaccord, d'accord, entièrement d'accord.</p>		<p>Q. 12-B Dans quelle mesure êtes-vous d'accord avec chacun des énoncés suivants ? Les mathématiques sont principalement une façon structurée de représenter le monde réel. CHOIX : Fortement en désaccord, en désaccord, d'accord, entièrement d'accord.</p>			
<p>Commentaires : Match identique pour 1995 et 1999 mais aucun item correspondant pour 2003</p>					

BTBMDATA

BTBMDATR

<p>Section B Q. 20-E Si vous donnez des devoirs de mathématiques, à quelle fréquence attribuez-vous aux élèves chacun des types de travaux suivants ? petites recherches ou collecte de données CHOIX : Jamais, rarement, parfois, toujours</p>		<p>Section B Q. 17-E Si vous donnez des devoirs mathématiques à vos élèves, vous arrive-t-il de leur attribuer les types de travaux suivants ? Les petites recherches ou une collecte de données CHOIX : jamais, rarement, parfois, toujours</p>			
<p>Commentaires : Match identique pour 1995 et 1999 mais aucun item correspondant pour 2003</p>					

BTBMIMP6

BTBMIMP6

<p>Q. 15-F Pour qu'ils réussissent en mathématiques à l'école, dans quelle mesure pensez-vous qu'il est important pour les élèves... De pouvoir fournir des raisons à l'appui de leurs solutions. CHOIX : pas important, plus ou moins important, très important.</p>		<p>Q. 11-F Pour qu'ils réussissent en mathématiques à l'école, dans quelle mesure pensez-vous qu'il est important pour les élèves... de pouvoir fournir des raisons à l'appui de leurs solutions CHOIX : pas important, plus ou moins important, très important.</p>			
---	--	--	--	--	--

Commentaires : Question identique pour 1995 et 1999 mais pas d'item correspondant pour 2003.

BTBMADHY

				<p>Q. 14-C Dans quelle mesure êtes-vous d'accord ou en désaccord avec les énoncés suivants? La résolution de problèmes mathématiques se fait souvent par hypothèses, estimations ou essais qui modifient nos conclusions. CHOIX : Tout à fait d'accord, d'accord, en désaccord, tout à fait en désaccord.</p>	
<p>Commentaires :</p>					

BTBMADDW

				<p>Q. 14-E Dans quelle mesure êtes-vous d'accord ou en désaccord avec les énoncés suivants? Il y a différentes façons de résoudre la plupart des problèmes mathématiques. CHOIX : Tout à fait d'accord, d'accord, en désaccord, tout à fait en désaccord.</p>	
<p>Commentaires :</p>					

BTBMPTOO

				<p>Q. 20-D Au cours d'une semaine type d'enseignement de mathématiques, quel pourcentage de temps les élèves de la classe TEIMS consacrent-ils aux activités suivantes ? Résoudre des problèmes seul.</p>	<p>BSBMHWPO Q. 10-K À quelle fréquence faites-vous les activités ci-dessous pendant les cours de mathématiques ? Nous travaillons seul à résoudre des problèmes</p>
--	--	--	--	---	---

BTBMCALI		BSBMOINF	
		Q. 31-C À quelle fréquence utilisez-vous l'ordinateur pour les activités ci-dessous, dans vos cours de mathématiques de la classe TEIMS? Permettre aux élèves de chercher des idées et de l'information. CHOIX : Presque toutes les cours, La moitié des cours, Quelques cours, Jamais.	Q. 14-C. A À quelle fréquence utilises-tu un ordinateur pour faire les choses suivantes ? je cherche des idées ou des renseignements au sujet des mathématiques CHOIX : Chaque jour, au moins une fois par semaine, une ou deux fois par mois, quelque fois par année, jamais.
Commentaires :			

BTBMCAPA		BSBGOPAD	
		Q. 31-D À quelle fréquence utilisez-vous l'ordinateur pour les activités ci-dessous, dans vos cours de mathématiques de la classe TEIMS? Permettre aux élèves de traiter et d'analyser des données. CHOIX : Presque toutes les cours, La moitié des cours, Quelques cours, Jamais.	Q. 14-C. D À quelle fréquence utilises-tu un ordinateur pour faire les choses suivantes ? Je traite et j'analyse des données CHOIX : Chaque jour, au moins une fois par semaine, une ou deux fois par mois, quelque fois par année, jamais.
Commentaires :			

BTBMTESP			
		Q. 39-B À quelle fréquence posez-vous les types de questions ci-dessous dans vos tests ou épreuves de mathématiques ? Questions impliquant la recherche de régularités et de relations. CHOIX : Presque jamais, parfois, presque toujours.	
Commentaires :			

Annexe D

Documents contenus dans la trousse acheminée au panel d'enseignants –

Démarches de recrutement auprès des participants

Lettre introduction aux directions des écoles

Lettre d'instructions aux enseignants

Fiche-item

Formulaire de consentement

Instrument questionnaire.

Suite à l'obtention des permissions du Bureau d'éthique en recherche de l'Université d'Ottawa, les démarches de recrutement servant à recueillir les données qualitatives ont été entamées auprès de trois conseils scolaires francophones de l'Ontario. Voici les étapes de la procédure suivie pour l'obtention des données qualitatives :

1. Évaluation du projet de recherche par les comités au sein des conseils scolaires : Les trois conseils ont procédé à une évaluation du projet par un de leurs comités.
2. Réception d'une autorisation provenant de chacun des trois conseils scolaires pour les opérations de recrutement sur leur territoire : Les trois conseils ciblés ont accepté que la recherche soit menée en recrutant des participants dans leurs écoles.
3. Création de la liste des écoles : Une liste des écoles ayant une 7^e et une 8^e années a été créée en consultant les sites Internet des trois conseils scolaires. Cette liste contient les noms, numéros de téléphone et adresses courriel des membres de directions ainsi que les noms et adresses civiques des écoles. La liste comporte un total de 36 écoles.
4. Première communication avec les écoles : Chaque directeur a tout d'abord reçu un courriel contenant une lettre d'introduction pour la direction de l'école, un résumé du projet de recherche ainsi qu'un feuillet de recrutement à remettre aux enseignants ciblés par l'étude.
5. Deuxième communication avec les écoles : Au moins un membre de la direction de chaque école fût joint par téléphone, ce qui a souvent exigé plusieurs tentatives. L'objectif était d'établir un contact avec les écoles, de s'assurer que l'information se soit bien rendue par courriel et que celle-ci soit présentée aux enseignants de 7^e et 8^e année.
6. Troisième communication : Plusieurs écoles ont demandé que la documentation leur soit réacheminée par courriel (souvent aux directeurs adjoints) ceci est fait rapidement.
7. Quatrième communication : Puisque plusieurs directeurs affirmaient qu'une entrevue serait plus compliquée qu'un questionnaire, le protocole d'entrevue fût envoyé par la poste à chacune des écoles, sous forme de questionnaire papier. La trousse comprenait une lettre d'introduction pour la direction de l'école, une lettre d'instructions pour le participant, une fiche item, un questionnaire, deux copies du formulaire de consentement ainsi qu'une enveloppe de retour adressée et pré-affranchie. Ces documents peuvent être consultés en annexe D.
8. Cinquième communication : Les réceptions/administrations scolaire et/ou les directions ont été rappelées afin de confirmer la réception des trousse et pour demander si la trousse avait été acheminée aux enseignants de mathématiques de la 7^e ou 8^e. Il était difficile d'obtenir des réponses claires à ce sujet même en parlant à la personne en charge de recevoir et distribuer le courrier.

Message adressé à la direction de l'établissement scolaire

La présente lettre a pour but de solliciter la participation d'enseignants oeuvrant au sein de votre établissement pour la réalisation d'un projet de recherche appuyé par la Faculté d'éducation de l'Université d'Ottawa. L'étude en cours porte sur la **qualité des données générées par les questionnaires contextuels administrés aux enseignants lors des évaluations de rendement à grande échelle**. La réalisation de ce projet de recherche compte fournir des réponses aux questions suivantes :

- Est-ce que les données disponibles provenant des évaluations à grande échelle (ex. TEIMS, PIRS, PISA, PIRLS) sont adéquates pour informer les décideurs au sujet de ce qui se passe véritablement dans les salles de classe ?
- Est-ce que le contenu des questionnaires est suffisamment représentatif des nouveaux programmes éducatifs des provinces canadiennes ?
- Est-ce que la qualité des instruments de collecte permet de faire confiance aux résultats obtenus lors des analyses ?
- Comment les questionnaires contextuels de ces programmes d'évaluation peuvent-ils être améliorés pour mieux répondre aux besoins d'information ?

Avant de communiquer avec les directions d'écoles, nos intentions ont été transmises directement à votre conseil scolaire et c'est avec leur approbation que nous faisons appel à vous maintenant. Soyez assuré que cette étude ne présente pas de risques véritables pour les participants, et qu'elle a fait l'objet d'une évaluation déontologique complète par le Comité d'éthique de la recherche de l'Université d'Ottawa.

Afin de recueillir les données nécessaires pour répondre aux questions de recherche ci-haut, des opérations de recrutement sont nécessaires. Pour se faire, chaque école a reçu une trousse contenant l'information nécessaire afin que le participant puisse compléter le questionnaire. Nous vous demandons de bien vouloir acheminer tout le contenu de la trousse à un **enseignant de mathématiques de la 7^e ou de la 8^e année** de votre établissement scolaire. Le recrutement des participants sera effectué sur une base volontaire et les personnes intéressées seront priées de compléter le questionnaire avant le **26 mai 2006**. Si vous nécessitez des renseignements additionnels, prière de communiquer par courriel à l'adresse suivante :

Veillez accepter, cher monsieur ou madame, mes salutations les plus distinguées, et sachez que votre collaboration dans cette initiative sera grandement appréciée.

Sarah Plouffe
Faculté d'éducation
Université d'Ottawa

Bonjour aux enseignants de mathématiques de 7^e et de 8^e année !

Nous avons besoin de votre expertise pour la réalisation d'un projet de recherche appuyé par la Faculté de l'éducation de l'Université d'Ottawa.

Peut-être avez-vous déjà entendu parlé des programmes d'évaluation à grande échelle, comme la TEIMS, le PIRLS, le PISA, ou encore le PIRS ? Saviez-vous que l'information recueillie par les questionnaires administrés aux enseignants lors de ces grandes enquêtes est souvent utilisée par les décideurs politiques et les directions d'écoles pour allouer des ressources ou encore élaborer des plans correctifs ? L'étude en cours sert à évaluer la qualité de l'information issue de l'administration de ces questionnaires.

Quoi : Comme participant, on vous demande de compléter un questionnaire pour recueillir vos commentaires. Il s'agira notamment de commenter sur la forme et la nature des questions présentées, comme par exemple le vocabulaire utilisé, le niveau de clarté de la question, ou encore la pertinence de la question pour votre pratique en tant qu'enseignant.

Comment : Il s'agit de suivre les instructions qui sont fournies dans la trousse. Le tout devrait exiger environ 30 minutes.

Quand : Avant le 26 mai 2006

Votre participation à cette étude est complètement volontaire et strictement confidentielle.

Pour de plus amples renseignements, prière de communiquer avec :

Sarah Plouffe, Concentration mesure et évaluation

Faculté d'éducation - Université d'Ottawa

Courriel :

Instructions

1. Lire le *formulaire de consentement*, remplir la case avec votre nom et apposer votre signature au bas du formulaire. Il y a deux copies du formulaire, une qui devra nous être renvoyée et l'autre que vous pourrez conserver pour vos dossiers.
2. Prendre le *questionnaire de l'enseignant* et répondre aux trois questions démographiques.
3. Prendre la *fiche item* et lire attentivement l'item 1 sur la fiche.
4. Répondre aux *pages 1, 2 et 3 du questionnaire de l'enseignant* en vous référant uniquement à l'item 1.
5. Lorsque terminé, prendre la *fiche item* de nouveau et lire attentivement l'item 2 sur la fiche.
6. Répondre aux *pages 4, 5 et 6 du questionnaire* en vous référant uniquement à l'item 2.
7. Lorsque le questionnaire est complété, placer les documents suivants dans l'enveloppe de retour fournie :
 - Le questionnaire de l'enseignant
 - Le formulaire de consentement avec votre signature

Vous avez terminé...!

Confiez vite l'enveloppe à la poste, vos réponses nous sont si précieuses!

Fiche item

Item 1

Pendant les cours de mathématiques, vous arrive-t-il de demander aux élèves de réaliser les activités suivantes ?

- Expliquer le raisonnement qui sous-tend une idée.

CHOIX DE RÉPONSE : jamais ou presque jamais
 parfois
 la plupart du temps
 à chaque fois

Item 2

À quelle fréquence demandez-vous aux élèves de mathématiques de la classe TEIMS de faire ce qui suit?

- Écrire des équations et des fonctions qui représentent des relations.

CHOIX DE RÉPONSE : tous les cours
 la moitié des cours
 quelques cours
 jamais

Formulaire de consentement

Titre du projet: Analyse d'items du questionnaire contextuel des enseignants de la TEIMS portant sur les pratiques pédagogiques

Chercheurs : Marielle Simon
Professeur
Faculté d'éducation
Université d'Ottawa
562-5800 poste 4145
msimon@uottawa.ca

Sarah Plouffe
Étudiante à la maîtrise
Faculté d'éducation
Université d'Ottawa

Invitation à participer:

Vous êtes invité(e) à participer à la recherche nommée ci-haut, qui est menée par Sarah Plouffe étudiante à la maîtrise à l'Université d'Ottawa et supervisée par Mme Marielle Simon, professeure titulaire de l'Université d'Ottawa.

But de l'étude:

Vous avez peut-être entendu parlé des programmes d'évaluation à grande échelle (par exemple, TEIMS, PIRLS, PISA, PIRS). Lors de la tenue de telles évaluations, des questionnaires contextuels sont administrés aux enseignants dans le but de recueillir de l'information sur leur façon d'enseigner, leurs caractéristiques démographiques, leurs croyances par rapport à l'enseignement, la situation dans leur école, etc. La présente étude sert à juger de la qualité de l'information qui est générée par ces questionnaires. L'étude en cours tentera de répondre à des questions telles que :

- Est-ce que les données disponibles provenant des évaluations à grande échelle sont adéquates pour informer les décideurs au sujet de ce qui se passe véritablement dans les salles de classe ?
- Est-ce que le contenu des questionnaires est suffisamment représentatif des nouveaux programmes éducatifs des provinces canadiennes ?
- Est-ce que la qualité des instruments de collecte permet de faire confiance aux résultats obtenus lors des analyses ?
- Comment les questionnaires contextuels de ces programmes d'évaluation peuvent-ils être améliorés pour mieux répondre aux besoins d'information ?

Participation:

Comme participant à l'étude, on vous demandera de compléter le questionnaire qui vous sera fourni. Il comporte des exemples d'items du questionnaire administré aux enseignants lors du programme d'évaluation TEIMS. Vous ne devrez pas répondre à ces questions mais plutôt fournir vos commentaires quant à des aspects précis des items du questionnaire. Il s'agira notamment de commenter sur la forme et la nature des questions présentées, comme par exemple le vocabulaire utilisé, le niveau de clarté de la question, ou encore la pertinence de la question pour votre pratique en tant qu'enseignant.

La participation de chaque enseignant est strictement confidentielle.

Risques:

Il est compris que puisqu'une participation à cette recherche implique la collecte de certains renseignements personnels tels que le groupe d'âge, le sexe et le nombre d'année d'expérience, il est possible qu'elle crée des inquiétudes. Les chercheuses ont prévu les mesures suivantes afin de minimiser ces risques :

- 1) Votre nom ne sera pas documenté, plutôt un numéro de référence vous sera alloué au début du processus.
- 2) Les renseignements personnels recueillis seront conservés dans le bureau de la chercheuse principale, verrouillées sous clé pour une durée de 5 ans. Seulement Marielle Simon et Sarah Plouffe auront accès à ces données.
- 3) Dans les publications, les renseignements personnels seront rapportés globalement pour l'ensemble du groupe afin qu'aucun participant ne puisse être identifié.
- 4) Le nom de l'école où vous travaillez ne sera jamais inclus dans les rapports rendus publics.

Bienfaits:

Une participation à cette recherche aura pour effet de générer de l'information précise pour mieux juger de la pertinence et de la validité des programmes d'évaluation pour la population minoritaire francophone de l'Ontario. Cette information ne peut être obtenue autrement qu'en sondant un échantillon d'enseignants de mathématiques de l'Ontario. Les recommandations formulées à partir des commentaires recueillis serviront à l'amélioration des questionnaires contextuels administrés dans le cadre de programmes d'évaluation à grande échelle d'envergure nationale et internationale. Ils permettront également d'identifier de meilleures façons d'interpréter les nombreuses données déjà amassées. Les praticiens, chercheurs, décideurs, enseignants et élèves pourront tous tirer profit des conclusions qui seront communiquées lors de la présentation du rapport final de cette étude.

Confidentialité et anonymat:

Les chercheurs soutiennent que l'information partagée demeurera strictement confidentielle. Le contenu ne sera utilisé que pour les buts précis de cette recherche et selon le respect de la confidentialité et de l'anonymat.

Participation volontaire:

Votre participation à la recherche est volontaire et vous êtes libre de vous retirer en tout temps, et/ou refuser de répondre à certaines questions, sans subir de conséquences négatives. Si vous choisissez de vous retirer de l'étude, les données recueillies jusqu'à ce moment seront éliminées et détruites de façon sécuritaire.

Acceptation:

Je, _____, accepte de participer à cette recherche menée par Sarah Plouffe de la Faculté d'éducation de l'Université d'Ottawa, laquelle recherche est supervisée par Marielle Simon, professeure titulaire de l'Université d'Ottawa. Pour tout renseignement additionnel concernant cette étude, vous pouvez communiquer avec le chercheur ou son superviseur.

Pour tout renseignement sur les aspects éthiques de cette recherche, vous pouvez vous adresser au Responsable de l'éthique en recherche, Université d'Ottawa, Pavillon Tabaret, 550, rue Cumberland, pièce 159, (613) 562-5841 ou ethics@uottawa.ca.

Il y a deux copies du formulaire de consentement, dont une copie que vous pouvez garder.

Signature du participant: _____

Date: _____

Signature du chercheur: _____

Date: _____

Questionnaire de l'enseignant

Questions démographiques	Sexe : Homme <input type="checkbox"/> Femme <input type="checkbox"/>
	Dans quel groupe d'âge êtes-vous : Moins de 25 ans <input type="checkbox"/> De 25 à 29 ans <input type="checkbox"/> De 30 à 39 ans <input type="checkbox"/> De 40 à 49 ans <input type="checkbox"/> De 50 à 59 ans <input type="checkbox"/> 60 ans et plus <input type="checkbox"/>
	Nombre d'années d'expérience : _____
Item 1	
Formulation de la question	1) Que pensez-vous de la formulation de la question ? 2) La question est-elle rédigée en respectant les règles de grammaire et de syntaxe ? 3) Est-ce que la longueur de la question est acceptable ?
Commentaires	
Niveau de familiarité du vocabulaire	1) Est-ce que le vocabulaire vous est familier ? 2) Est-ce que vous utilisez ce vocabulaire dans le cadre de votre travail ? 3) Quel sens accordez-vous aux termes importants de la question ? 4) Selon vous, est-ce que ces termes possèdent une définition généralement bien établie et uniforme auprès de vos collègues ?
Commentaires	

Clarté de la question	1) Est-ce que cette question est facilement comprise ? 2) Est-ce que l'information demandée est claire pour vous ? 3) Est-ce que vous croyez que la question pourrait être interprétée de différentes façons ?
Commentaires	
Présence de plus d'un concept dans la question	1) Est-ce que la question cible plus d'un concept ? Si oui, selon vous lesquels ? 2) Est-ce que la présence de plus d'un concept agit sur la clarté de la question ?
Commentaires	
Réponses biaisées	1) Est-ce que vous pouvez déceler un sens moral à cette question ? 2) Seriez-vous tenté de répondre d'une façon donnée si vous sachiez que votre directeur ou vos collègues avaient accès à vos réponses ? 3) Selon vous, est-ce que votre école, votre conseil ou le ministère ont véhiculé des messages précis en lien avec le sujet abordé par la question?
Commentaires	

Les choix de réponse	1) Est-ce que vous trouvez que les choix de réponse sont adéquats ? 2) Est-ce que les choix de réponse vous permettent de traduire votre réalité de façon fidèle ? 3) Est-ce que les choix de réponse vous permettent de répondre de façon suffisamment précise ?
-----------------------------	---

Commentaires

Pertinence pour le contexte de réforme de l'Ontario	1) D'où proviennent vos connaissances en lien avec le sujet de la question (cours universitaires, atelier à l'école, documents du ministère, etc.) ? 2) Est-ce que vous croyez que le sujet abordé par la question est important pour le contexte scolaire des francophones de l'Ontario ? 3) Pensez-vous que le sujet de la question pourrait avoir un lien avec le rendement des élèves ?
--	---

Commentaires

Item 2	
Formulation de la question	1) Que pensez-vous de la formulation de la question ? 2) La question est-elle rédigée en respectant les règles de grammaire et de syntaxe ? 3) Est-ce que la longueur de la question est acceptable ?
Commentaires	
Niveau de familiarité du vocabulaire	1) Est-ce que le vocabulaire vous est familier ? 2) Est-ce que vous utilisez ce vocabulaire dans le cadre de votre travail ? 3) Quel sens accordez-vous aux termes importants de la question ? 4) Selon vous, est-ce que ces termes possèdent une définition généralement bien établie et uniforme auprès de vos collègues ?
Commentaires	
Clarté de la question	1) Est-ce que cette question est facilement comprise ? 2) Est-ce que l'information demandée est claire pour vous ? 3) Est-ce que vous croyez que la question pourrait être interprétée de différentes façons ?
Commentaires	

Présence de plus d'un concept dans la question	1) Est-ce que la question cible plus d'un concept ? Si oui, selon vous lesquels ? 2) Est-ce que la présence de plus d'un concept agit sur la clarté de la question ?
Commentaires	
Réponses biaisées	1) Est-ce que vous pouvez déceler un sens moral à cette question ? 2) Seriez-vous tenté de répondre d'une façon donnée si vous sachiez que votre directeur ou vos collègues avaient accès à vos réponses ? 3) Selon vous, est-ce que votre école, votre conseil ou le ministère ont véhiculé des messages précis en lien avec le sujet abordé par la question?
Commentaires	
Les choix de réponse	1) Est-ce que vous trouvez que les choix de réponse sont adéquats ? 2) Est-ce que les choix de réponse vous permettent de traduire votre réalité de façon fidèle ? 3) Est-ce que les choix de réponse vous permettent de répondre de façon suffisamment précise ?
Commentaires	

Pertinence pour le contexte de réforme de l'Ontario	1) D'où proviennent vos connaissances en lien avec le sujet de la question (cours universitaires, atelier à l'école, documents du ministère, etc.) ? 2) Est-ce que vous croyez que le sujet abordé par la question est important pour le contexte scolaire des francophones de l'Ontario ? 3) Pensez-vous que le sujet de la question pourrait avoir un lien avec le rendement des élèves ?
Commentaires	

Nous vous remercions sincèrement de votre contribution à ce projet !

Annexe E

- 1. Moyenne des items de chaque cycle**
- 2. Résultats des analyses de variance (ANOVA) pour les items de l'étude (a priori)**
- 3. Résultats du test d'homogénéité des variances (test Levene)**
- 4. Résultats des tests robustes Welch, Brown-Forsythe**
- 5. Résultats des analyses non-paramétriques équivalentes (Kruskal-Wallis)**
- 6. Résultats des comparaisons ANOVA a posteriori (LSD)**

Moyenne des items pour chaque cycle

Échantillon	N	Moyenne	Écarts types	Erreur type	Intervalles de confiance (moyenne)		Minimum	Maximum	
					Limite inférieure	Limite supérieure			
					Item 1	1995,00			123
	1999,00	90	2,54	,501	,053	2,44	2,65	2	3
	2003,00	76	1,62	,610	,070	1,48	1,76	1	3
Item 2	1995,00	122	3,75	,437	,040	3,67	3,82	3	4
	1999,00	89	3,54	,523	,055	3,43	3,65	2	4
	2003,00	76	3,50	,529	,061	3,38	3,62	2	4
Item 3	1995,00	116	2,33	,732	,068	2,19	2,46	1	4
	1999,00	85	2,31	,618	,067	2,17	2,44	1	4
	2003,00	73	2,68	,724	,085	2,52	2,85	1	4
Item 4	1995,00	123	2,87	,640	,058	2,76	2,98	1	4
	1999,00	89	2,88	,671	,071	2,74	3,02	2	4
	2003,00	76	3,45	,661	,076	3,30	3,60	2	4
Item 5	1995,00	123	4,52	1,059	,095	4,33	4,71	1	5
	1999,00	90	2,26	,464	,049	2,16	2,35	1	3
	2003,00	75	2,25	,522	,060	2,13	2,37	1	3
Item 6	1995,00	122	1,96	,673	,061	1,84	2,08	1	4
	1999,00	90	2,16	,422	,044	2,07	2,24	1	4
	2003,00	76	2,11	,349	,040	2,03	2,19	1	3
Item 7	1995,00	123	3,16	,961	,087	2,99	3,33	1	4
	1999,00	87	3,53	,745	,080	3,37	3,69	1	4
	2003,00	70	3,30	,787	,094	3,11	3,49	2	4
Item 8	1995,00	122	2,50	,874	,079	2,34	2,66	1	4
	1999,00	90	2,41	,652	,069	2,27	2,55	1	4
	2003,00	75	2,20	,520	,060	2,08	2,32	1	4
Item 9	1995,00	123	1,98	,677	,061	1,86	2,10	1	4
	1999,00	89	2,07	,654	,069	1,93	2,21	1	4
	2003,00	76	2,49	,721	,083	2,32	2,65	1	4
Item 10	1995,00	123	1,79	,547	,049	1,69	1,89	1	4
	1999,00	90	1,92	,674	,071	1,78	2,06	1	4
	2003,00	76	2,80	,674	,077	2,65	2,96	1	4
Item 11	1995,00	123	1,46	,577	,052	1,36	1,57	1	3
	1999,00	90	1,34	,523	,055	1,23	1,45	1	3
	2003,00	36	1,81	,467	,078	1,65	1,96	1	3
Item 12	1995,00	123	2,93	,248	,022	2,89	2,98	2	3
	1999,00	90	2,74	,464	,049	2,65	2,84	1	3
	2003,00	76	2,83	,773	,089	2,65	3,01	1	4

Item 13	1995,00	119	3,23	,528	,048	3,13	3,32	2	4
	1999,00	85	3,00	,512	,056	2,89	3,11	2	4
	2003,00	76	3,24	,651	,075	3,09	3,39	1	4
Item 14	1995,00	122	2,80	1,050	,095	2,62	2,99	1	4
	1999,00	86	3,15	,988	,107	2,94	3,36	1	4
	2003,00	69	2,87	1,013	,122	2,63	3,11	1	4
Item 15	1995,00	120	2,09	,830	,076	1,94	2,24	1	4
	1999,00	90	2,02	,948	,100	1,82	2,22	1	4
	2003,00	43	1,56	,796	,121	1,31	1,80	1	3
Item 16	1995,00	110	2,33	,910	,087	2,16	2,50	1	4
	1999,00	89	2,24	,798	,085	2,07	2,40	1	3
	2003,00	74	2,11	,632	,073	1,96	2,25	1	3
Item 17	1995,00	118	2,43	,779	,072	2,29	2,57	1	4
	1999,00	90	2,48	,810	,085	2,31	2,65	1	4
	2003,00	74	2,74	,440	,051	2,64	2,85	2	3

Résultats des analyses de variance (ANOVA) pour les items de l'étude (a priori)

	Source	dl	Somme des Carrés	Carrés Moyens	F	Taille de l'effet (η^2)
Item 1	Traitement	2	37,236	18,618	55,875 **	0,281
	Erreur	286	95,297	0,333		
	Total	288	132,533			
Item 2	Traitement	2	3,604	1,802	7,501 **	0,050
	Erreur	284	68,235	0,240		
	Total	286	71,840			
Item 3	Traitement	2	7,217	3,609	7,445 **	0,052
	Erreur	271	131,352	0,485		
	Total	273	138,569			
Item 4	Traitement	2	18,481	9,241	21,525 **	0,131
	Erreur	285	122,349	0,429		
	Total	287	140,830			
Item 5	Traitement	2	361,770	180,885	292,897 **	0,673
	Erreur	285	176,008	0,618		

	Total	287	537,778			
Item 6	Traitement	2	2,221	1,111	3,968 *	0,027
	Erreur	285	79,775	0,280		
	Total	287	81,997			
Item 7	Traitement	2	6,842	3,421	4,665 *	0,033
	Erreur	277	203,126	0,733		
	Total	279	209,968			
Item 8	Traitement	2	4,220	2,110	3,987 *	0,027
	Erreur	284	150,289	0,529		
	Total	286	154,509			
Item 9	Traitement	2	12,613	6,307	13,560 **	0,087
	Erreur	285	132,550	0,465		
	Total	287	145,163			
Item 10	Traitement	2	52,288	26,144	67,363 **	0,320
	Erreur	286	110,999	0,388		
	Total	288	163,287			
Item 11	Traitement	2	5,478	2,739	9,287 **	0,070
	Erreur	246	72,546	0,295		
	Total	248	78,024			
Item 12	Traitement	2	1,923	0,961	3,852 *	0,026
	Erreur	286	71,378	0,250		
	Total	288	73,301			
Item 13	Traitement	2	3,157	1,579	5,048 **	0,035
	Erreur	277	86,611	0,313		
	Total	279	89,768			
Item 14	Traitement	2	6,416	3,208	3,072 *	0,022
	Erreur	274	286,140	1,044		
	Total	276	292,556			
Item 15	Traitement	2	6,743	3,372	4,745 **	0,034
	Erreur	269	191,125	0,711		
	Total	271	197,868			

Item 16	Traitement	2	3,239	1,620	2,325	0,017
	Erreur	272	189,430	0,696		
	Total	274	192,669			
<hr/>						
Item 17	Traitement	2	4,738	2,369	4,605 *	0,032
	Erreur	279	143,535	0,514		
	Total	281	148,273			

Note. * $p < ,05$ et ** $p < 0,01$

Résultats du test d'homogénéité des variances (test Levene)

	Statistique Levene	dl1	dl2	Sign.
ITEM1	1,196	2	286	,304
ITEM2	21,934	2	284	,000
ITEM3	1,735	2	271	,178
ITEM4	2,809	2	285	,062
ITEM5	14,322	2	285	,000
ITEM6	7,035	2	285	,001
ITEM7	4,049	2	277	,018
ITEM8	22,188	2	284	,000
ITEM9	5,988	2	285	,003
ITEM10	1,696	2	286	,185
ITEM11	8,609	2	246	,000
ITEM12	72,197	2	286	,000
ITEM13	8,968	2	277	,000
ITEM14	,330	2	274	,719
ITEM15	,514	2	250	,599
ITEM16	14,841	2	270	,000
ITEM17	19,748	2	279	,000

Note : les items en surbrillance dans ce tableau possèdent possiblement des variances problématiques, i.e. non homogènes entre les populations de l'étude.

Résultats des tests robustes de l'égalité des moyennes Welch et Brown-Forsythe

		Statistique (a)	dl1	dl2	Sign.
ITEM1	Welch	56,161	2	175,145	,000
	Brown-Forsythe	56,480	2	254,403	,000
ITEM2	Welch	7,850	2	164,470	,001
	Brown-Forsythe	7,141	2	237,923	,001
ITEM3	Welch	7,170	2	167,545	,001
	Brown-Forsythe	7,546	2	246,727	,001
ITEM4	Welch	21,215	2	172,096	,000
	Brown-Forsythe	21,304	2	256,645	,000
ITEM5	Welch	242,376	2	183,479	,000
	Brown-Forsythe	364,904	2	235,535	,000
ITEM6	Welch	3,440	2	189,632	,034
	Brown-Forsythe	4,695	2	266,919	,010
ITEM7	Welch	4,958	2	171,867	,008
	Brown-Forsythe	5,014	2	265,169	,007
ITEM8	Welch	5,277	2	187,477	,006
	Brown-Forsythe	4,535	2	280,315	,012
ITEM9	Welch	12,536	2	172,088	,000
	Brown-Forsythe	13,409	2	252,190	,000
ITEM10	Welch	62,802	2	164,053	,000
	Brown-Forsythe	63,815	2	236,831	,000
ITEM11	Welch	11,737	2	103,805	,000
	Brown-Forsythe	10,344	2	178,570	,000
ITEM12	Welch	6,579	2	131,606	,002
	Brown-Forsythe	3,086	2	136,858	,049
ITEM13	Welch	5,568	2	164,374	,005
	Brown-Forsythe	4,850	2	225,248	,009
ITEM14	Welch	3,149	2	163,650	,046
	Brown-Forsythe	3,113	2	246,664	,046
ITEM15	Welch	7,139	2	116,205	,001
	Brown-Forsythe	6,344	2	187,922	,002
ITEM16	Welch	1,916	2	178,207	,150
	Brown-Forsythe	1,750	2	268,426	,176
ITEM17	Welch	7,638	2	179,993	,001
	Brown-Forsythe	5,027	2	247,543	,007

Résultats des analyses non-paramétriques Kruskal-Wallis

	GROUPE	N	Moyennes	Test Kruskal Wallis		
				chi-carré	dl	sign.
ITEM1	1	123	154.593	76.35	2	0.000
	2	90	184.194			
	3	76	83.059			
	Total		289			
ITEM2	1	122	161.791	14.04	2	0.001
	2	89	133.348			
	3	76	127.914			
	Total		287			
ITEM3	1	116	128.621	14.71	2	0.001
	2	85	125.900			
	3	73	165.116			
	Total		274			
ITEM4	1	123	128.024	38.74	2	0.000
	2	89	127.702			
	3	76	190.836			
	Total		288			
ITEM5	1	123	214.390	171.20	2	0.000
	2	90	92.211			
	3	75	92.627			
	Total		288			
ITEM6	1	122	131.152	9.84	2	0.007
	2	90	156.700			
	3	76	151.480			
	Total		288			
ITEM7	1	123	129.561	8.64	2	0.013
	2	87	159.224			
	3	70	136.450			
	Total		280			
ITEM8	1	122	155.664	8.87	2	0.012
	2	90	145.728			
	3	75	122.953			
	Total		287			
ITEM9	1	123	128.927	25.11	2	0.000
	2	89	135.702			
	3	76	180.007			
	Total		288			
ITEM10	1	123	113.911	93.77	2	0.000
	2	90	127.533			
	3	76	216.000			
	Total		289			
ITEM11	1	123	123.659	20.10	2	0.000
	2	90	110.539			
	3	36	165.736			
	Total		249			
ITEM12	1	123	157.187	8.41	2	0.015

		2	90	132.556			
		3	76	140.013			
	Total		289				
ITEM13		1	119	147.504	11.60	2	0.003
		2	85	120.171			
		3	76	152.270			
	Total		280				
ITEM14		1	122	129.762	6.43	2	0.040
		2	86	156.169			
		3	69	133.935			
	Total		277				
ITEM15		1	120	147.692	9.04	2	0.011
		2	90	137.689			
		3	62	113.113			
	Total		272				
ITEM16		1	112	147.094	4.48	2	0.106
		2	89	138.663			
		3	74	123.439			
	Total		275				
ITEM17		1	118	130.788	11.30	2	0.004
		2	90	135.150			
		3	74	166.304			
	Total		282				

Résultats des comparaisons ANOVA a posteriori (LSD)

Variable dépendante	Groupe (i)	Groupe (j)	Différence moyenne (I-J)	Erreur type	Sign.	Intervalle de confiance à 95%	
						Borne inférieure	Borne supérieure
Item 1	1995.00	1999.00	-.260(*)	.080	.001	-.42	-.10
		2003.00	.666(*)	.084	.000	.50	.83
	1999.00	1995.00	.260(*)	.080	.001	.10	.42
		2003.00	.926(*)	.090	.000	.75	1.10
	2003.00	1995.00	-.666(*)	.084	.000	-.83	-.50
		1999.00	-.926(*)	.090	.000	-1.10	-.75
Item 2	1995.00	1999.00	.207(*)	.068	.003	.07	.34
		2003.00	.246(*)	.072	.001	.10	.39
	1999.00	1995.00	-.207(*)	.068	.003	-.34	-.07
		2003.00	.039	.077	.608	-.11	.19
	2003.00	1995.00	-.246(*)	.072	.001	-.39	-.10
		1999.00	-.039	.077	.608	-.19	.11
Item 3	1995.00	1999.00	.022	.099	.827	-.17	.22
		2003.00	-.357(*)	.104	.001	-.56	-.15
	1999.00	1995.00	-.022	.099	.827	-.22	.17
		2003.00	-.379(*)	.111	.001	-.60	-.16
	2003.00	1995.00	.357(*)	.104	.001	.15	.56
		1999.00	.379(*)	.111	.001	.16	.60
Item 4	1995.00	1999.00	-.006	.091	.943	-.19	.17
		2003.00	-.577(*)	.096	.000	-.77	-.39
	1999.00	1995.00	.006	.091	.943	-.17	.19
		2003.00	-.571(*)	.102	.000	-.77	-.37
	2003.00	1995.00	.577(*)	.096	.000	.39	.77
		1999.00	.571(*)	.102	.000	.37	.77
Item 5	1995.00	1999.00	2.265(*)	.109	.000	2.05	2.48
		2003.00	2.267(*)	.115	.000	2.04	2.49
	1999.00	1995.00	-2.265(*)	.109	.000	-2.48	-2.05
		2003.00	.002	.123	.986	-.24	.24
	2003.00	1995.00	-2.267(*)	.115	.000	-2.49	-2.04
		1999.00	-.002	.123	.986	-.24	.24
Item 6	1995.00	1999.00	-.197(*)	.074	.008	-.34	-.05
		2003.00	-.146	.077	.060	-.30	.01
	1999.00	1995.00	.197(*)	.074	.008	.05	.34
		2003.00	.050	.082	.542	-.11	.21
	2003.00	1995.00	.146	.077	.060	-.01	.30
		1999.00	-.050	.082	.542	-.21	.11
Item 7	1995.00	1999.00	-.366(*)	.120	.002	-.60	-.13
		2003.00	-.137	.128	.285	-.39	.11
	1999.00	1995.00	.366(*)	.120	.002	.13	.60
		2003.00	.229	.137	.097	-.04	.50
	2003.00	1995.00	.137	.128	.285	-.11	.39
		1999.00	-.229	.137	.097	-.50	.04
Item 8	1995.00	1999.00	.089	.101	.380	-.11	.29
		2003.00	.300(*)	.107	.005	.09	.51

	1999.00	1995.00	-.089	.101	.380	-.29	.11
		2003.00	.211	.114	.064	-.01	.43
	2003.00	1995.00	-.300(*)	.107	.005	-.51	-.09
		1999.00	-.211	.114	.064	-.43	.01
Item 9	1995.00	1999.00	-.084	.095	.379	-.27	.10
		2003.00	-.503(*)	.100	.000	-.70	-.31
	1999.00	1995.00	.084	.095	.379	-.10	.27
		2003.00	-.419(*)	.107	.000	-.63	-.21
	2003.00	1995.00	.503(*)	.100	.000	.31	.70
		1999.00	.419(*)	.107	.000	.21	.63
Item 10	1995.00	1999.00	-.134	.086	.123	-.30	.04
		2003.00	-1.014(*)	.091	.000	-1.19	-.84
	1999.00	1995.00	.134	.086	.123	-.04	.30
		2003.00	-.880(*)	.097	.000	-1.07	-.69
	2003.00	1995.00	1.014(*)	.091	.000	.84	1.19
		1999.00	.880(*)	.097	.000	.69	1.07
Item 11	1995.00	1999.00	.119	.075	.116	-.03	.27
		2003.00	-.342(*)	.103	.001	-.54	-.14
	1999.00	1995.00	-.119	.075	.116	-.27	.03
		2003.00	-.461(*)	.107	.000	-.67	-.25
	2003.00	1995.00	.342(*)	.103	.001	.14	.54
		1999.00	.461(*)	.107	.000	.25	.67
Item 12	1995.00	1999.00	.191(*)	.069	.006	.05	.33
		2003.00	.106	.073	.147	-.04	.25
	1999.00	1995.00	-.191(*)	.069	.006	-.33	-.05
		2003.00	-.085	.078	.278	-.24	.07
	2003.00	1995.00	-.106	.073	.147	-.25	.04
		1999.00	.085	.078	.278	-.07	.24
Item 13	1995.00	1999.00	.227(*)	.079	.005	.07	.38
		2003.00	-.010	.082	.904	-.17	.15
	1999.00	1995.00	-.227(*)	.079	.005	-.38	-.07
		2003.00	-.237(*)	.088	.008	-.41	-.06
	2003.00	1995.00	.010	.082	.904	-.15	.17
		1999.00	.237(*)	.088	.008	.06	.41
Item 14	1995.00	1999.00	-.348(*)	.144	.016	-.63	-.06
		2003.00	-.066	.154	.667	-.37	.24
	1999.00	1995.00	.348(*)	.144	.016	.06	.63
		2003.00	.282	.165	.089	-.04	.61
	2003.00	1995.00	.066	.154	.667	-.24	.37
		1999.00	-.282	.165	.089	-.61	.04
Item 15	1995.00	1999.00	.069	.121	.567	-.17	.31
		2003.00	.534(*)	.154	.001	.23	.84
	1999.00	1995.00	-.069	.121	.567	-.31	.17
		2003.00	.464(*)	.161	.004	.15	.78
	2003.00	1995.00	-.534(*)	.154	.001	-.84	-.23
		1999.00	-.464(*)	.161	.004	-.78	-.15
Item 16	1995.00	1999.00	.091	.115	.428	-.13	.32
		2003.00	.219	.121	.072	-.02	.46
	1999.00	1995.00	-.091	.115	.428	-.32	.13
		2003.00	.128	.127	.314	-.12	.38

Item 17	2003.00	1995.00	-.219	.121	.072	-.46	.02
		1999.00	-.128	.127	.314	-.38	.12
	1995.00	1999.00	-.046	.100	.650	-.24	.15
		2003.00	-.311(*)	.106	.004	-.52	-.10
	1999.00	1995.00	.046	.100	.650	-.15	.24
		2003.00	-.265(*)	.113	.019	-.49	-.04
	2003.00	1995.00	.311(*)	.106	.004	.10	.52
		1999.00	.265(*)	.113	.019	.04	.49

* La différence des moyennes est significative à 05.

Annexe F

Résultats des analyses factorielles exploratoires (Phase II – 1995, 1999 et 2003)

Résultats des analyses factorielles exploratoires, échantillon de 1995, deuxième vague

	Facteur 1	Facteur 2	Facteur 3	Facteur 4	Facteur 5
BTBMAGR2 – (ITEM 21) D'accord, les maths sont une façon de représenter le monde réel.	0,632				
BTBMAGR3 – (ITEM 13) D'accord, maths sont un guide pour aborder les situations réelles.	1,014				
BTBMAGR7 – (ITEM 3) D'accord, les maths comme un ensemble d'algorithmes qui couvrent toutes les possibilités.				0,758	
BTBMASK2 – (ITEM 6) fréquence, élèves présentent et analysent des tableaux ou des graphiques.			0,791		
BTBMASK3 – (ITEM 10) fréquence, élèves travaillent sur des problèmes n'ayant aucune solution évidente			0,570		
BTBMASK4 – (ITEM 11) fréquence, se servir d'un ordinateur pour résoudre des problèmes ou faire des exercices.			0,401		
BTBMASK5 – (ITEM 8) fréquence, Indiquer des équations pour représenter des rapports.			0,504		
BTBMCAL4 - (ITEM 7) utiliser la calculatrice pour		-0,801			
BTBMCAL5 – (ITEM 14) utiliser la calculatrice pour explorer les notions numériques		-0,894			
BTBMCALC – (ITEM 5) accès à la calculatrice pendant les cours		-0,534			
BTBMDATA – (ITEM 22) devoirs, petites recherches ou collecte de données					0,646
BTBMFINN – (ITEM 16) devoirs, détermination d'une ou plusieurs applications pour la matière étudiée.					0,616
BTBMIMP1 – (ITEM 1) Important de mémoriser des formules et des méthodes?				0,674	
BTBMIMP2 – (ITEM 18) Important de penser de façon ordonnée et séquentielle.				0,793	
BTBMIMP4 – (ITEM 20) Important de penser de façon créatrice					0,607
BTBMIMP5 – (ITEM 12) Important de comprendre l'application des mathématiques dans la vie courante					0,883
BTBMIMP6 – (ITEM 23) Important de pouvoir fournir des raisons à l'appui de leurs solutions.					0,484
BTBMLES6 – (ITEM 9) fréquence, travail de groupe			0,512		

Note : Il est important de noter que dans le présent tableau les noms originaux des variables, fournis par la TEIMS, ont été conservés puisqu'ils permettent une identification rapide des items issus du même énoncé général.

Résultats des analyses factorielles exploratoires, échantillon de 1999, deuxième vague

	Facteur 1	Facteur 2	Facteur 3	Facteur 4	Facteur 5
BTBMAGR3 – D'accord, maths sont un guide pour aborder les situations réelles.	0,543				
BTBMAGR7 – (ITEM 3) D'accord, les maths comme un ensemble d'algorithmes qui couvrent toutes les possibilités.	0,607				
BTBMIMP1 – (ITEM 1) Important de mémoriser des formules et des méthodes?	0,366				
BTBMLM08 - Le manque de matériel informatique	-0,519				
BTBMAGR6 - (ITEM 2) D'accord, faire appel à plus d'un mode de représentation (illustration, matériel concret, etc.)		0,538			
BTBMIMP4 – Important de penser de façon créatrice		0,770			
BTBMIMP5 – Important de comprendre l'application des mathématiques dans la vie courante		0,602			
BTBMIMP6 – Important de pouvoir fournir des raisons à l'appui de leurs solutions.		0,501			
BTBMDATR – devoirs, petites recherches ou collecte de données			-0,704		
BTBMFINR – devoirs, détermination d'une ou plusieurs applications pour la matière étudiée.			-0,817		
BTBMIMP3 – Important de comprendre les notions, les principes et les méthodes des mathématiques.			0,595		
BTBMWGT2 - Questionnaires exigeant des réponses brèves pour expliquer leur raisonnement.				0,549	
BTBMASK1 – (ITEM 4) Fréquence, élèves expliquent le raisonnement qui sous-tend une idée.				0,871	
BTBMASK4 – fréquence, se servir d'un ordinateur pour résoudre des problèmes ou faire des exercices.				0,420	
BTBMCAL4 - utiliser la calculatrice pour Répondre des problèmes complexes					0,951
BTBMCAL5 – utiliser la calculatrice pour explorer les notions numériques					0,765
BTBMWECP – accès à la calculatrice pendant les cours					0,451

Note : Il est important de noter que dans le présent tableau les noms originaux des variables, fournis par la TEIMS, ont été conservés puisqu'ils permettent une identification rapide des items issus du même énoncé général.

Résultats des analyses factorielles exploratoires, échantillon de 2003, deuxième vague

	Facteur 1	Facteur 2	Facteur 3	Facteur 4
BTBMADME - Apprendre les mathématiques est surtout une question de mémoire.	0.545			
BTBMADSA – D'accord, Les mathématiques comme ensembles d'algorithmes ou de règles couvrant toutes les possibilités	0.460			
BTBMASID – fréquence, Interpréter des données de tableaux ou de diagrammes	0.507			
BTBMASRR – fréquence, Écrire des équations et des fonctions qui représentent des relations	0.651			
BTBMADHY – D'accord, la résolution de problèmes mathématiques par hypothèses, estimations ou essais	0.750			
BTBMKHCA – fréquence, devoirs, Trouver une ou plusieurs applications des notions abordées		-0.454		
BTBMADDW – D'accord, différentes façons de résoudre la plupart des problèmes mathématiques		0.644		
BTBMASCP – fréquence, déterminer leur propre façon de résoudre un problème complexe			-0.547	
BTBMASEA – fréquence, Expliquer leurs réponses			-0.639	
BTBMASWS – fréquence, Résoudre des problèmes dont la solution n'est pas immédiatement perceptible			-0.872	
BTBMASG – fréquence, Travailler en petits groupes				0.489
BTBMTEJU - fréquence, devoirs, Questions exigeant des explications ou des justifications				0.997
BTBMTESP – fréquences, tests, questions impliquant la recherche de régularités et de relations				0.489

Note : Il est important de noter que dans le présent tableau les noms originaux des variables, fournis par la TEIMS, ont été conservés puisqu'ils permettent une identification rapide des items issus du même énoncé général.

Annexe G

**Tableaux des taux de réponse obtenus pour les items enseignants
pour les trois cycles d'enquête 1995, 1999 et 2003**

Tableau 1

Taux de réponse aux items sélectionnés issus des questionnaires administrés aux enseignants lors de la TEIMS 1995

1995	échantillon final n=123				échantillon original des enseignants francophones de l'Ontario n=161			
	valides	manquantes	% valides	% manquantes	valides	manquantes	% valides	% manquantes
BTBMAGR2	119	4	96.75	3.25	145	16	90.06	9.94
BTBMAGR3	119	4	96.75	3.25	144	17	89.44	10.56
BTBMAGR6	122	1	99.19	0.81	148	13	91.93	8.07
BTBMAGR7	116	7	94.31	5.69	139	22	86.34	13.66
BTBMASK1	123	0	100.00	0.00	123	38	76.40	23.60
BTBMASK2	122	1	99.19	0.81	122	39	75.78	24.22
BTBMASK3	123	0	100.00	0.00	123	38	76.40	23.60
BTBMASK4	123	0	100.00	0.00	123	38	76.40	23.60
BTBMASK5	122	1	99.19	0.81	122	39	75.78	24.22
BTBMCAL4	123	0	100.00	0.00	123	38	76.40	23.60
BTBMCAL5	122	1	99.19	0.81	122	39	75.78	24.22
BTBMCALC	123	0	100.00	0.00	123	38	76.40	23.60
BTBMDATA	116	7	94.31	5.69	116	45	72.05	27.95
BTBMFIND	112	11	91.06	8.94	112	49	69.57	30.43
BTBMIMP1	123	0	100.00	0.00	149	12	92.55	7.45
BTBMIMP2	123	0	100.00	0.00	149	12	92.55	7.45
BTBMIMP3	123	0	100.00	0.00	149	12	92.55	7.45
BTBMIMP4	123	0	100.00	0.00	149	12	92.55	7.45
BTBMIMP5	123	0	100.00	0.00	149	12	92.55	7.45
BTBMIMP6	123	0	100.00	0.00	149	12	92.55	7.45
BTBMLES6	123	0	100.00	0.00	123	38	76.40	23.60
BTBMLM08	120	3	97.56	2.44	120	41	74.53	25.47
BTBMWGT2	118	5	95.93	4.07	118	43	73.29	26.71

Tableau 2

Taux de réponse aux items sélectionnés issus des questionnaires administrés aux enseignants lors de la TEIMS 1999

	échantillon final				échantillon original des enseignants francophones de l'ontario			
	valides	manquants	% valides	% manquants	valides	manquants	% valides	% manquants
BTBMAGR2	84	6	93.33	4.88	84	36	70.00	30.00
BTBMAGR3	85	5	94.44	4.07	85	35	70.83	29.17
BTBMAGR6	89	1	98.89	0.81	89	31	74.17	25.83
BTBMAGR7	85	5	94.44	4.07	85	35	70.83	29.17
BTBMASK1	89	1	98.89	0.81	89	31	74.17	25.83
BTBMASK2	90	0	100.00	0.00	90	30	75.00	25.00
BTBMASK3	90	0	100.00	0.00	90	30	75.00	25.00
BTBMASK4	90	0	100.00	0.00	90	30	75.00	25.00
BTBMASK5	90	0	100.00	0.00	90	30	75.00	25.00
BTBMCAL4	87	3	96.67	2.44	87	33	72.50	27.50
BTBMCAL5	86	4	95.56	3.25	86	34	71.67	28.33
BTBMDATR	89	1	98.89	0.81	89	31	74.17	25.83
BTBMFINR	89	1	98.89	0.81	89	31	74.17	25.83
BTBMIMP1	90	0	100.00	0.00	90	30	75.00	25.00
BTBMIMP2	90	0	100.00	0.00	90	30	75.00	25.00
BTBMIMP3	90	0	100.00	0.00	90	30	75.00	25.00
BTBMIMP4	90	0	100.00	0.00	90	30	75.00	25.00
BTBMIMP5	90	0	100.00	0.00	90	30	75.00	25.00
BTBMIMP6	90	0	100.00	0.00	90	30	75.00	25.00
BTBMLES5	89	1	98.89	0.81	89	31	74.17	25.83
BTBMLM08	90	0	100.00	0.00	90	30	75.00	25.00
BTBMWECP	90	0	100.00	0.00	90	30	75.00	25.00
BTBMWGT2	90	0	100.00	0.00	90	30	75.00	25.00

Tableau 3

Taux de réponse aux items sélectionnés issus des questionnaires administrés aux enseignants lors de la TEIMS 2003

	échantillon final				échantillon original des enseignants francophones de l'ontario			
	valides	manquantes	% valides	% manquantes	valides	manquantes	% valides	% manquantes
BTBGLT07	74	1	98.67	1.33	74	8	90.24	9.76
BTBMADDW	75	0	100.00	0.00	76	6	92.68	7.32
BTBMADHY	75	0	100.00	0.00	76	6	92.68	7.32
BTBMADME	75	0	100.00	0.00	76	6	92.68	7.32
BTBMADMR	75	0	100.00	0.00	76	6	92.68	7.32
BTBMADRW	75	0	100.00	0.00	75	7	91.46	8.54
BTBMADSA	72	3	96.00	4.00	73	9	89.02	10.98
BTBMASCP	74	1	98.67	1.33	75	7	91.46	8.54
BTBMASDL	74	1	98.67	1.33	75	7	91.46	8.54
BTBMASEA	74	1	98.67	1.33	75	7	91.46	8.54
BTBMASID	74	1	98.67	1.33	75	7	91.46	8.54
BTBMASRR	73	2	97.33	2.67	74	8	90.24	9.76
BTBMASRG	74	1	98.67	1.33	75	7	91.46	8.54
BTBMASWS	74	1	98.67	1.33	75	7	91.46	8.54
BTBMCADM	35	40	46.67	53.33	35	47	42.68	57.32
BTBMCALB	68	7	90.67	9.33	68	14	82.93	17.07
BTBMCALI	36	39	48.00	52.00	36	46	43.90	56.10
BTBMCALS	69	6	92.00	8.00	69	13	84.15	15.85
BTBMCAML	74	1	98.67	1.33	74	8	90.24	9.76
BTBMCAPA	36	39	48.00	52.00	36	46	43.90	56.10
BTBMCASP	36	39	48.00	52.00	36	46	43.90	56.10
BTBMCOMA	75	0	100.00	0.00	75	7	91.46	8.54
BTBMKHCA	73	2	97.33	2.67	73	9	89.02	10.98
BTBMPTOO	75	0	100.00	0.00	76	6	92.68	7.32
BTBMTEJU	73	2	97.33	2.67	73	9	89.02	10.98
BTBMTESP	72	3	96.00	4.00	72	10	87.80	12.20

Annexe H

Tableaux des coefficients d'asymétrie et de kurtose pour les items de l'étude selon les trois cycles d'enquête 1995, 1999 et 2003.

Coefficients d'asymétrie et de kurtose, erreurs type, cote Z et intervalles de confiance pour les variables enseignants de 1995.

Items	Asymétrie	Erreur-type d'asymétrie	cote Z asymétrie	Limite inférieure intervalle de confiance	Limite supérieure intervalle de confiance	Kurtose	Erreur-type de kurtose	cote Z de kurtose	Limite inférieure intervalle de confiance	Limite supérieure intervalle de confiance
Item 1	-0,23	0,218	-1,07	-0,66	0,19	-0,58	0,433	-1,34	-1,43	0,27
Item 2	-1,14	0,219	-5,22	-1,57	-0,71	-0,70	0,435	-1,62	-1,56	0,15
Item 3	0,21	0,225	0,95	-0,23	0,65	-0,12	0,446	-0,26	-0,99	0,76
Item 4	-0,26	0,218	-1,20	-0,69	0,17	0,35	0,433	0,81	-0,50	1,20
Item 5	-2,27	0,218	-10,39	-2,69	-1,84	4,07	0,433	9,41	3,23	4,92
Item 6	0,54	0,219	2,48	0,11	0,97	0,89	0,435	2,05	0,04	1,74
Item 7	-0,84	0,218	-3,85	-1,27	-0,41	-0,41	0,433	-0,95	-1,26	0,44
Item 8	-0,04	0,219	-0,17	-0,47	0,39	-0,66	0,435	-1,52	-1,51	0,19
Item 9	0,34	0,218	1,56	-0,09	0,77	0,23	0,433	0,53	-0,62	1,08
Item 10	0,22	0,218	1,01	-0,21	0,65	1,50	0,433	3,47	0,65	2,35
Item 11	0,80	0,218	3,65	0,37	1,22	-0,35	0,433	-0,81	-1,20	0,50
Item 12	-3,57	0,218	-16,37	-4,00	-3,14	10,93	0,433	25,24	10,08	11,78
Item 13	0,19	0,222	0,85	-0,25	0,62	-0,14	0,440	-0,31	-1,00	0,73
Item 14	-0,38	0,219	-1,74	-0,81	0,05	-1,06	0,435	-2,44	-1,91	-0,21
Item 15	0,36	0,221	1,65	-0,07	0,80	-0,43	0,438	-0,99	-1,29	0,42
Item 16	0,15	0,228	0,65	-0,30	0,60	-0,41	0,453	-0,91	-1,30	0,47
Item 17	0,06	0,223	0,29	-0,37	0,50	-0,35	0,442	-0,80	-1,22	0,51
Item 18	-3,87	0,218	-17,75	-4,30	-3,44	13,21	0,433	30,50	12,36	14,06
Item 19	-4,71	0,218	-21,58	-5,14	-4,28	20,52	0,433	47,36	19,67	21,36
Item 20	-1,33	0,218	-6,08	-1,75	-0,90	0,79	0,433	1,82	-0,06	1,64
Item 21	-0,44	0,222	-1,98	-0,87	-0,01	1,19	0,440	2,70	0,33	2,05
Item 22	0,21	0,225	0,94	-0,23	0,65	0,69	0,446	1,55	-0,18	1,56
Item 23	-3,06	0,218	-14,02	-3,49	-2,63	9,32	0,433	21,52	8,47	10,17

Coefficients d'asymétrie et de kurtose, erreurs type, cote Z et intervalles de confiance pour les variables enseignants de 1999.

Items	Asymétrie	Erreur-type d'asymétrie	cote Z asymétrie	Limite inférieure intervalle de confiance	Limite supérieure intervalle de confiance	Kurtose	Erreur-type de kurtose	cote Z de kurtose	Limite inférieure intervalle de confiance	Limite supérieure intervalle de confiance
Item 1	-0,18	0,254	-0,71	-0,68	0,32	-2,01	0,503	-4,00	-3,00	-1,03
Item 2	-0,40	0,255	-1,58	-0,90	0,10	-1,30	0,506	-2,58	-2,29	-0,31
Item 3	0,32	0,261	1,22	-0,19	0,83	0,20	0,517	0,40	-0,81	1,22
Item 4	0,15	0,255	0,58	-0,35	0,65	-0,75	0,506	-1,49	-1,74	0,24
Item 5	0,79	0,254	3,11	0,29	1,29	-0,47	0,503	-0,93	-1,45	0,52
Item 6	1,88	0,254	7,41	1,38	2,38	4,51	0,503	8,96	3,52	5,49
Item 7	-1,40	0,258	-5,41	-1,90	-0,89	0,94	0,511	1,84	-0,06	1,94
Item 8	0,59	0,254	2,31	0,09	1,09	0,15	0,503	0,31	-0,83	1,14
Item 9	0,93	0,255	3,65	0,43	1,43	2,16	0,506	4,27	1,17	3,15
Item 10	1,22	0,254	4,79	0,72	1,71	1,34	0,503	2,66	0,35	2,32
Item 11	0,54	0,254	2,14	0,05	1,04	0,85	0,503	1,70	-0,13	1,84
Item 12	1,14	0,254	4,49	0,64	1,64	0,25	0,503	0,51	-0,73	1,24
Item 13	-1,48	0,254	-5,83	-1,98	-0,98	1,06	0,503	2,11	0,08	2,05
Item 14	0,00	0,261	0,00	-0,51	0,51	0,99	0,517	1,92	-0,02	2,00
Item 15	-0,84	0,260	-3,22	-1,34	-0,33	-0,49	0,514	-0,95	-1,50	0,52
Item 16	0,68	0,254	2,69	0,19	1,18	-0,37	0,503	-0,74	-1,36	0,61
Item 17	-0,45	0,255	-1,78	-0,95	0,05	-1,28	0,506	-2,54	-2,27	-0,29
Item 18	0,07	0,254	0,29	-0,42	0,57	-0,44	0,503	-0,87	-1,42	0,55
Item 19	-2,19	0,254	-8,64	-2,69	-1,70	2,88	0,503	5,72	1,89	3,86
Item 20	-2,35	0,254	-9,24	-2,84	-1,85	3,58	0,503	7,12	2,60	4,57
Item 21	-0,94	0,254	-3,69	-1,44	-0,44	-0,48	0,503	-0,96	-1,47	0,50
Item 22	-0,58	0,263	-2,22	-1,10	-0,07	1,25	0,520	2,40	0,23	2,27
Item 23	-0,78	0,255	-3,05	-1,28	-0,28	-0,40	0,506	-0,79	-1,39	0,59

Coefficients d'asymétrie et de kurtose, erreurs type, cote Z et intervalles de confiance pour les variables enseignants de 2003.

Items	Asymétrie	Erreur-type d'asymétrie	cote Z asymétrie	Limite inférieure intervalle de confiance	Limite supérieure intervalle de confiance	Kurtose	Erreur-type de kurtose	cote Z de kurtose	Limite inférieure intervalle de confiance	Limite supérieure intervalle de confiance
Item 1	-0,41	0,274	-1,50	-0,95	0,13	-0,63	0,541	-1,17	-1,69	0,43
Item 2	0,25	0,274	0,90	-0,29	0,78	-1,35	0,541	-2,49	-2,41	-0,29
Item 3	0,09	0,279	0,31	-0,46	0,63	-0,19	0,552	-0,35	-1,27	0,89
Item 4	0,80	0,276	2,89	0,26	1,34	-0,43	0,545	-0,79	-1,50	0,64
Item 5	-0,26	0,276	-0,93	-0,80	0,28	-0,25	0,545	-0,47	-1,32	0,81
Item 6	-1,61	0,276	-5,84	-2,15	-1,07	3,92	0,545	7,19	2,85	4,99
Item 7	0,61	0,285	2,16	0,06	1,17	-1,11	0,563	-1,97	-2,21	0,00
Item 8	-0,83	0,277	-2,99	-1,37	-0,29	1,56	0,548	2,85	0,49	2,64
Item 9	-0,49	0,276	-1,76	-1,03	0,05	-0,16	0,545	-0,30	-1,23	0,91
Item 10	0,01	0,276	0,04	-0,53	0,55	-0,27	0,545	-0,49	-1,34	0,80
Item 11	0,71	0,388	1,84	-0,05	1,47	0,70	0,759	0,92	-0,79	2,19
Item 12	-0,13	0,276	-0,47	-0,67	0,41	-0,93	0,545	-1,71	-2,00	0,13
Item 13	0,86	0,274	3,15	0,33	1,40	2,21	0,541	4,08	1,15	3,27
Item 14	0,26	0,287	0,92	-0,30	0,82	-1,18	0,566	-2,08	-2,29	-0,07
Item 15	0,58	0,276	2,11	0,04	1,12	-0,18	0,545	-0,33	-1,25	0,89
Item 16	0,10	0,277	0,37	-0,44	0,65	-0,50	0,548	-0,91	-1,57	0,58
Item 17	1,16	0,277	4,17	0,61	1,70	-0,68	0,548	-1,24	-1,75	0,40
Item 24	0,42	0,274	1,55	-0,11	0,96	0,17	0,541	0,31	-0,89	1,23
Item 25	1,41	0,274	5,16	0,88	1,95	2,27	0,541	4,20	1,21	3,33
Item 26	1,06	0,274	3,85	0,52	1,59	1,34	0,541	2,47	0,28	2,40
Item 27	0,10	0,276	0,37	-0,44	0,64	-1,00	0,545	-1,84	-2,07	0,07
Item 28	-0,08	0,274	-0,29	-0,62	0,46	-2,05	0,541	-3,78	-3,11	-0,99
Item 29	1,61	0,393	4,11	0,84	2,38	0,63	0,768	0,82	-0,87	2,14
Item 30	-0,05	0,388	-0,13	-0,81	0,71	-0,23	0,759	-0,30	-1,71	1,26
Item 31	0,41	0,388	1,05	-0,35	1,16	1,18	0,759	1,56	-0,31	2,67
Item 32	-0,07	0,279	-0,27	-0,62	0,47	0,28	0,552	0,50	-0,80	1,36

Annexe I

Tests de distribution normale (Kolmogorov-Smirnov et Shapiro-Wilk)

Tests de normalité						
Données 1995						
	Kolmogorov-Smirnov				Shapiro-Wilk	
	Statistique	dl	Sign.	Statistique	dl	Sign.
BTBMAGR2	0.33702	94	0.000	0.759993	94	0.000
BTBMAGR3	0.386174	94	0.000	0.70168	94	0.000
BTBMAGR6	0.489559	94	0.000	0.492028	94	0.000
BTBMAGR7	0.305461	94	0.000	0.828533	94	0.000
BTBMASK1	0.35066	94	0.000	0.781056	94	0.000
BTBMASK2	0.322109	94	0.000	0.788124	94	0.000
BTBMASK3	0.379855	94	0.000	0.692526	94	0.000
BTBMASK4	0.377906	94	0.000	0.681372	94	0.000
BTBMASK5	0.230289	94	0.000	0.876817	94	0.000
BTBMCAL4	0.27437	94	0.000	0.801484	94	0.000
BTBMCAL5	0.213676	94	0.000	0.84915	94	0.000
BTBMCALC	0.441308	94	0.000	0.544928	94	0.000
BTBMDATA	0.266412	94	0.000	0.809456	94	0.000
BTBMFIND	0.238717	94	0.000	0.870558	94	0.000
BTBMIMP1	0.319101	94	0.000	0.763532	94	0.000
BTBMIMP2	0.538632	94	0.000	0.262391	94	0.000
BTBMIMP3	0.539744	94	0.000	0.168171	94	0.000
BTBMIMP4	0.449474	94	0.000	0.584153	94	0.000
BTBMIMP5	0.536614	94	0.000	0.288206	94	0.000
BTBMIMP6	0.519304	94	0.000	0.382849	94	0.000
BTBMLES5	0.303193	94	0.000	0.784197	94	0.000
BTBMLM08	0.249114	94	0.000	0.863927	94	0.000
BTBMWGT2	0.248212	94	0.000	0.858506	94	0.000
conclusion : Tous les tests réalisés démontre que les données ne sont pas distribuées normalement.						
données 1999						
	Kolmogorov-Smirnov				Shapiro-Wilk	
	Statistique	dl	Sign.	Statistique	dl	Sign.
BTBMIMP1	0.362937	90	0.000	0.633538	90	0.000
BTBMAGR6	0.36115	89	0.000	0.669429	89	0.000
BTBMAGR7	0.348602	85	0.000	0.775268	85	0.000
BTBMASK1	0.280918	89	0.000	0.794971	89	0.000
BTBMWECP	0.442629	90	0.000	0.604659	90	0.000

BTBMASK2	0.488355	90	0.000	0.507351	90	0.000
BTBMCAL4	0.403276	87	0.000	0.661011	87	0.000
BTBMASK5	0.347064	90	0.000	0.774922	90	0.000
BTBMLES5	0.383771	89	0.000	0.722902	89	0.000
BTBMLES6	0.416741	90	0.000	0.679298	90	0.000
BTBMASK3	0.309635	90	0.000	0.787198	90	0.000
BTBMASK4	0.422794	90	0.000	0.628644	90	0.000
BTBMIMP5	0.464852	90	0.000	0.556398	90	0.000
BTBMAGR3	0.370588	85	0.000	0.68751	85	0.000
BTBMCAL5	0.293153	86	0.000	0.785874	86	0.000
BTBMLM08	0.253797	90	0.000	0.833529	90	0.000
BTBMFINR	0.291491	89	0.000	0.770552	89	0.000
BTBMWGT2	0.24447	90	0.000	0.864702	90	0.000
BTBMIMP2	0.518416	90	0.000	0.401097	90	0.000
BTBMIMP3	0.522485	90	0.000	0.382871	90	0.000
BTBMIMP4	0.420869	90	0.000	0.624111	90	0.000
BTBMAGR2	0.403064	84	0.000	0.703065	84	0.000
BTBMDATR	0.339406	89	0.000	0.72897	89	0.000
BTBMIMP6	0.487291	90	0.000	0.497385	90	0.000

conclusion : Tous les tests réalisés démontre que les données ne sont pas distribuées normalement.

Données 2003						
	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistique	dl	Sign.	Statistique	dl	Sign.
BTBMADME	0.291915	76	0.000	0.746201	76	0.000
BTBMADMR	0.340806	76	0.000	0.679498	76	0.000
BTBMADSA	0.284695	73	0.000	0.837418	73	0.000
BTBMASEA	0.337838	76	0.000	0.732884	76	0.000
BTBMCAML	0.392842	75	0.000	0.689558	75	0.000
BTBMASID	0.499962	76	0.000	0.450162	76	0.000
BTBMCALS	0.313222	70	0.000	0.755523	70	0.000
BTBMASRR	0.423106	75	0.000	0.67064	75	0.000
BTBMASSG	0.316029	76	0.000	0.810046	76	0.000
BTBMASWS	0.299436	76	0.000	0.813399	76	0.000
BTBMCASP	0.43915	36	0.000	0.622929	36	0.000
BTBMASDL	0.226749	76	0.000	0.832227	76	0.000
BTBMADRW	0.313157	76	0.000	0.728874	76	0.000
BTBMCAL5	0.230013	69	0.000	0.838516	69	0.000
BTBGLT07	0.254067	70	0.000	0.807649	70	0.000
BTBMKHCA	0.31118	74	0.000	0.779605	74	0.000
BTBMTEJU	0.463555	74	0.000	0.544091	74	0.000
BTBMADHY	0.295351	76	0.000	0.818766	76	0.000
BTBMADDW	0.375903	76	0.000	0.676852	76	0.000
BTBMPTOO	0.177905	76	0.000	0.898743	76	0.000

BTBMASCP	0.216173	76	0.000	0.855596	76	0.000
BTBMCOMA	0.353327	76	0.000	0.63546	76	0.000
BTBMCADM	0.488925	35	0.000	0.491425	35	0.000
BTBMCALI	0.341005	36	0.000	0.751633	36	0.000
BTBMCAPA	0.417777	36	0.000	0.656412	36	0.000
BTBMTESP	0.369285	73	0.000	0.723553	73	0.000
conclusion : Tous les tests réalisés démontre que les données ne sont pas distribuées normalement.						

Annexe J

Tableaux des mesures de dispersion : étendue, écarts types et variance

Tableau des mesures de dispersion pour les données de l'échantillon de 1995 – de la TEIMS

1995						
	Écarts types	Variance	Étendue	Maximum	Minimum	C.V.
Item 1	0.6076065	0.369186	2	3	1	26.60
Item 2	0.43714845	0.191099	1	4	3	11.67
Item 3	0.731595779	0.535232	3	4	1	31.43
Item 4	0.639663713	0.40917	3	4	1	22.29
Item 5	1.058529706	1.120485	4	5	1	23.42
Item 6	0.672942729	0.452852	3	4	1	34.35
Item 7	0.961334316	0.924164	3	4	1	30.40
Item 8	0.874335639	0.764463	3	4	1	34.97
Item 9	0.677310736	0.45875	3	4	1	34.14
Item 10	0.547004248	0.299214	3	4	1	30.58
Item 11	0.576772861	0.332667	2	3	1	39.41
Item 12	0.247606159	0.061309	1	3	2	8.44
Item 13	0.527818902	0.278593	2	4	2	16.36
Item 14	1.049512632	1.101477	3	4	1	37.44
Item 15	0.830063613	0.689006	3	4	1	39.68
Item 16	0.969117738	0.939189	4	5	1	40.80
Item 17	0.778765334	0.606475	3	4	1	32.02
Item 18	0.232619192	0.054112	1	3	2	7.90
Item 19	0.198286677	0.039318	1	3	2	6.70
Item 20	0.517197257	0.267493	2	3	1	19.28
Item 21	0.609706419	0.371742	3	4	1	20.55
Item 22	0.804550402	0.647301	4	5	1	34.95
Item 23	0.34363658	0.118086	2	3	1	11.91

Tableau des mesures de dispersion pour les données de l'échantillon de 1999 – de la TEIMS

1999						
	Écarts types	Variance	Étendue	Maximum	Minimum	C.V.
Item 1	0.500810828	0.250811	1	2	3	19.68
Item 2	0.523453996	0.274004	2	2	4	14.79
Item 3	0.617667067	0.381513	3	1	4	26.79
Item 4	0.671162911	0.45046	2	2	4	23.33
Item 5	0.463526083	0.214856	2	1	3	20.55
Item 6	0.421637021	0.177778	3	1	4	19.56
Item 7	0.744578552	0.554397	3	1	4	21.10
Item 8	0.651608976	0.424594	3	1	4	27.03
Item 9	0.611581932	0.374032	3	1	4	31.62
Item 10	0.674208279	0.454557	3	1	4	26.46
Item 11	0.522765144	0.273283	2	1	3	35.07
Item 12	0.463526083	0.214856	2	1	3	38.88
Item 13	0.511766316	0.261905	2	2	4	16.89

Item 14	0.988372898	0.976881	3	1	4	17.06
Item 15	0.947827531	0.898377	3	1	4	31.37
Item 16	0.798044085	0.636874	2	1	3	46.87
Item 17	0.810434448	0.656804	3	1	4	35.69
Item 18	0.341839045	0.116854	1	2	3	32.71
Item 19	0.329377273	0.108489	1	2	3	11.92
Item 20	0.500810828	0.250811	2	1	3	11.45
Item 21	0.541321771	0.293029	3	1	4	18.86
Item 22	0.640599502	0.410368	2	1	3	19.19
Item 23	0.461907244	0.213358	2	1	3	26.03

Tableau des mesures de dispersion pour les données de l'échantillon de 2003 – de la TEIMS

2003	Écart types	Variance	Étendue	Maximum	Minimum	C.V.
Item 1	0.610291849	0.372456	2	3	1	37.71
Item 2	0.529150262	0.28	2	4	2	15.12
Item 3	0.72412231	0.524353	3	4	1	26.97
Item 4	0.661205704	0.437193	2	4	2	19.18
Item 5	0.522295695	0.272793	2	3	1	23.18
Item 6	0.349435635	0.122105	2	3	1	16.60
Item 7	0.78666421	0.618841	2	4	2	23.84
Item 8	0.519875245	0.27027	3	4	1	23.63
Item 9	0.7209886	0.519825	3	4	1	28.99
Item 10	0.673691064	0.45386	3	4	1	24.04
Item 11	0.467176592	0.218254	2	3	1	25.87
Item 12	0.772669104	0.597018	3	4	1	27.31
Item 13	0.650505876	0.423158	3	4	1	20.10
Item 14	1.013338157	1.026854	3	4	1	35.31
Item 15	0.984454742	0.969151	3	4	1	50.30
Item 16	0.631752673	0.399111	2	3	1	29.97
Item 17	0.439825956	0.193447	1	3	2	16.03
Item 24	0.7209886	0.519825	3	4	1	23.93
Item 25	0.640723276	0.410526	3	4	1	18.04
Item 26	9.877530766	97.56561	47	50	3	54.32
Item 27	0.900194911	0.810351	3	4	1	31.82
Item 28	0.50262469	0.252632	1	2	1	34.11
Item 29	0.405839725	0.164706	1	2	1	22.55
Item 30	0.59093684	0.349206	2	3	1	33.24
Item 31	0.487136108	0.237302	2	3	1	26.17
Item 32	0.551392315	0.304033	2	3	1	25.97