

*NOTE SUR LA RECHERCHE ET LES MÉTHODES /
RESEARCH AND PRACTICE NOTE*

LA COMPARABILITÉ DES ÉCHANTILLONS
DANS LES ENQUÊTES DE L'INTERNATIONAL
ASSOCIATION FOR THE EVALUATION OF
EDUCATIONAL ACHIEVEMENT

Jean-Guy Blais
Université de Montréal
Montréal, Québec

Résumé : Cette note sur la recherche et les méthodes vise à mettre en perspective les résultats des élèves lors de l'enquête TIMSS 2003 de l'International Association for the Evaluation of Educational Achievement, particulièrement au plan de l'échantillonnage des écoles et des élèves participants. De ce fait, le regard critique peut être étendu à d'autres enquêtes du même type dont l'objet principal ne porte pas sur les mathématiques.

Abstract: This research and practice note is intended to offer a perspective on interpreting student results from the International Association for the Evaluation of Educational Achievement TIMSS 2003 survey, specifically focussing on the sampling plan of participating schools and students. The observations can therefore be extended to other surveys of that kind, where the focus is not on mathematics.

■ Cette note sur la recherche et les méthodes vise à mettre en perspective les résultats observés suite à une enquête menée sous l'égide de l'IEA, l'enquête TIMSS 2003 (Martin, Mullis, & Chrostowski, 2004; Mullis, Martin, Gonzalez, & Chrostowski, 2004) pour les mathématiques auprès des élèves de la quatrième année du primaire. Plus précisément, elle s'attarde au plan d'échantillonnage des écoles et des élèves participants. De ce fait, le regard critique porté peut être étendue à d'autres enquêtes du même type dont l'objet principal n'est pas les mathématiques.

Correspondance à l'auteur : Jean-Guy Blais, Département d'administration et fondements de l'éducation, Faculté des sciences de l'éducation, Université de Montréal, CP 6128 Succursale Centre-ville, Montréal, QC, H3C 3J7; <jean-guy.blais@umontreal.ca>

Les enquêtes internationales de l'IEA visent à améliorer la compréhension des systèmes d'éducation et à aider à mieux comprendre les causes des résultats et des différences observés entre les élèves. Les participants à ces enquêtes sont le plus souvent des pays, mais on y retrouve également ce qu'on appelle des juridictions, c'est-à-dire des entités politiquement constituées qui ont la responsabilité légale de leur système d'éducation. Il en est ainsi pour l'enquête TIMSS 2003 des provinces de l'Ontario et de Québec, de même que de l'état de l'Indiana.¹ Les enquêtes TIMSS pourraient donc permettre, en principe, de situer la performance d'un échantillon d'élèves d'un pays ou d'une juridiction (PJ) donnée à des épreuves uniformes en fonction de la performance d'échantillons d'élèves d'autres PJ et permettre un retour sur les programmes, les curriculums, les opportunités d'apprentissage, et les séquences d'apprentissage.

L'enquête TIMSS 2003 est une enquête qui vise à récolter des données pour permettre « le suivi du niveau général de connaissance ou de compétences des élèves ». Étant donné cette visée, le recours pour l'enquête TIMSS à un échantillon probabiliste d'écoles et d'élèves apparaît comme l'approche la plus efficace et la plus efficiente (en ce sens que sous certaines conditions techniques on peut obtenir des données assez précises, généralisables à la population, avec un rapport coûts/bénéfices intéressant). Cependant, au-delà de l'adéquation technique des procédures d'échantillonnage, des problèmes subsistent quant à l'équivalence des échantillons des PJ participants.

Pour l'enquête TIMSS 2003 auprès d'élèves du primaire, 26 PJ participent aux opérations. La population visée dans chacun de ces PJ est celle de tous les élèves inscrits au niveau qui contient le plus grand nombre d'élèves de neuf ans au moment de l'épreuve. Pour la grande majorité des PJ participants, ces élèves se retrouvent à la quatrième année du primaire. Cependant, pour certains PJ ces élèves sont plutôt en cinquième année du primaire (Angleterre, Écosse, Australie, et Nouvelle-Zélande). De plus, étant donné qu'il s'agit d'un échantillonnage en grappe en deux étapes où des classes complètes sont sélectionnées, plusieurs élèves de dix ans se retrouvent également dans certains des échantillons.

La pratique de la reprise forcée d'une année scolaire (i.e., le redoublement) n'étant pas la même pour chaque PJ, il est possible que certaines classes de quatrième année accueillent un nombre non négligeable d'élèves de dix et onze ans, alors que des élèves de neuf ans sont retenus en deuxième ou troisième année (voir Brown, 1999, au sujet des différences entre les pratiques de pays européens comme

l'Angleterre, la Suisse, et l'Allemagne, en ce qui concerne le redoublement et leurs impacts possibles sur les résultats moyens des élèves dans des enquêtes comme TIMSS).

À titre d'exemple, le Tableau 1 présente la moyenne de l'âge des élèves de quelques PJ participants. On y observe que l'âge moyen des élèves de l'échantillon du Québec est de 10,1, alors qu'il est de 11,1 pour la Lettonie, 10,9 pour la Lituanie, 10,6 pour la Russie, 10,3 pour l'Angleterre, et 9,7 pour l'Écosse. Pour ces deux derniers participants toutefois, les élèves de cet âge ont complété cinq années de scolarité.

À n'en pas douter, certaines particularités systémiques des différents PJ comme le fait d'avoir complété cinq années de scolarité plutôt que quatre années, de compter plus d'élèves âgés dans une classe, donc plus matures (ce qui n'est pas négligeable en quatrième année du primaire), et de promouvoir moins d'élèves à cause des difficultés d'apprentissage peuvent produire un portrait difficile à comparer avec d'autres portraits où ces pratiques n'ont pas cours car elles contribuent en quelque sorte à biaiser les comparaisons entre les PJ.

Tableau 1
Illustration de la différence entre « Âge moyen » pour certains PJ

Pays/ Juridiction	Âge moyen
Québec	10,1
Ontario	9,8
Lettonie	11,1
Lituanie	10,9
Russie	10,6
Angleterre	10,3
Écosse	9,7

Dans tous les PJ participants, les échantillons d'élèves sont constitués selon les principes d'un échantillonnage en grappe en deux étapes. Selon cette approche, la première étape consiste pour chaque PJ à choisir des écoles selon la technique probabiliste du choix proportionnel à la taille (après une mise en rang selon la taille, des écoles sont choisies selon une séquence déterminée, une sur dix par exemple, afin de s'assurer d'obtenir des écoles dans différentes catégories de taille d'école). La deuxième étape consiste à choisir au hasard et selon la taille de l'école une classe d'élèves dans chacune

des écoles préalablement sélectionnées lors de la première étape.² Ce plan d'échantillonnage peut être qualifié de plan de base. Le choix des écoles s'effectue d'abord selon la taille de l'école pour assurer la présence d'élèves provenant d'écoles de tailles différentes et ensuite le choix d'une classe complète d'élèves est privilégié pour perturber le moins possible les activités régulières d'enseignement.

Pour différentes raisons propres à chacun des PJ, d'autres variables sont aussi utilisées pour ajouter un ou deux éléments de stratification. Ainsi, au Québec comme en Ontario et en Nouvelle-Zélande, un niveau supplémentaire est ajouté et l'échantillon est stratifié au départ pour tenir compte de la langue d'enseignement. Ensuite la méthode en deux étapes est appliquée pour chacune des strates. D'autres PJ ont opté pour des variables de stratification différentes. Ainsi par exemple, la Russie et l'Australie ont décidé de stratifier leur échantillon en fonction des régions, la Lituanie et la Slovaquie selon la taille de l'école, le Japon et le Yémen selon la dimension rural/urbain, l'Indiana selon le réseau, l'Iran et les États-Unis selon le niveau de pauvreté.³ D'un point de vue technique, ces différences quant au plan d'échantillonnage ne posent pas de problème puisqu'il est possible d'en tenir compte pour produire les estimations des moyennes et des écart-types des scores pour chaque PJ.

L'objectif final de l'opération d'échantillonnage est d'obtenir pour chaque PJ un échantillon d'au moins 150 écoles et d'au moins 4 000 élèves. Par exemple, pour le Québec, un échantillon⁴ de 193 écoles et 4 350 élèves est constitué à partir d'une population de 1 879 écoles et de 98 326 élèves. Comme la procédure d'échantillonnage exige dans un premier temps la sélection des écoles et qu'il est possible que certaines d'entre elles refusent de participer, des écoles de remplacement sont aussi sélectionnées pour pallier à ces refus. Ainsi, un autre indice, à part l'âge des élèves, pour l'étude de la comparabilité des échantillons est celui de la participation des écoles au premier tour (avant remplacement). Pour la province de Québec, on note une participation de 97 %, ce qui est très élevé et dénote une très bonne collaboration du milieu. Cependant, dans d'autres PJ plusieurs écoles n'ont pas démontré, semble-t-il, le même empressement à collaborer. Ainsi, le taux de participation des écoles a été de 52 %, 54 %, 64 %, et 70 % respectivement pour les Pays-Bas, l'Angleterre, l'Écosse, et les États-Unis. Il faudrait donc s'interroger sur les motifs qui amènent un aussi grand nombre d'écoles de ces PJ à ne pas collaborer et tenter dans l'analyse des données de documenter l'impact de ces refus sur les résultats observés.

En résumé, on peut dire que la question de l'âge des élèves et celle de la faible collaboration des écoles au premier tour mettent en perspective les difficultés que les responsables ont rencontrées dans la mise en place d'échantillons assez équivalents pour permettre des comparaisons raisonnables entre les résultats des PJ. À la lumière des propos de Brown (1999) sur le sujet et de ce qui est écrit dans les rapports TIMSS, des réserves peuvent être émises quant au succès de cette opération pour l'enquête 2003 en mathématiques.

Il faut donc être prudent dans les comparaisons entre PJ sur la base des échantillons constitués et mettre en perspective les limites des valeurs des indices de précision comme l'erreur standard d'estimation et les intervalles de confiance qui en découlent, car ceux-ci tiennent la route pour faire des comparaisons seulement si les échantillons sont comparables, ce que ne confirment pas les informations disponibles au sujet des élèves réellement échantillonnés.

Même si les différents échantillons ne sont pas complètement comparables techniquement, les données récoltées sont importantes et uniques. Il revient à chaque PJ participant de les exploiter en les situant adéquatement dans leur contexte en fonction du curriculum réellement abordé (plutôt que le curriculum officiel, i.e., celui au programme) et en fonction des PJ avec des systèmes semblables. Il convient toutefois d'éviter les comparaisons directes des scores moyens et la mise en rangs des PJ sans prendre en compte les différences « culturelles » qui existent quant au cheminement des élèves dans les systèmes d'éducation respectifs.

Il s'agit cependant du lot de ce type d'enquête, chaque PJ y retrouve d'autres PJ « comparables », mais y retrouve également des entités auxquelles il ne peut en toute bonne foi se comparer. Les modèles de mesure, les statistiques, et les calculs ne font pas la différence entre ce qui est comparable et ce qui ne l'est pas. Cette tâche revient aux responsables des systèmes d'éducation de chacun des PJ. Une comparaison adéquate demande plus qu'un regard sur le score moyen ou sur le rang relatif, elle demande un regard sur les items et sur le lien des items avec le programme et avec ce qui a été enseigné.

NOTES

1. On peut remarquer cependant que l'Écosse, qui n'est pas officiellement un « pays », voit ses résultats présentés dans la catégorie des pays et non des juridictions comme le Québec ou l'Ontario.

2. Dans certains PJ, on a opté pour deux classes par école, alors que d'autres PJ ont été obligés de choisir plusieurs classes dans une même école pour répondre à l'exigence d'un minimum de 4 000 élèves dans l'échantillon (exigence qui n'a pas toujours été rencontrée). De plus, dans certains PJ où les élèves dans les classes sont très nombreux, un échantillon d'élèves de la classe a été prélevé.
3. On peut noter que les échantillons n'ont pas été stratifiés selon le genre, mais que des comparaisons *a posteriori* garçons-filles sont présentées dans le rapport TIMSS. La précision des estimations des scores moyens pour les garçons et les filles pourrait donc ne pas être fondée techniquement.
4. Pour différentes raisons, certains élèves retenus dans l'échantillon n'ont pas participé aux épreuves. Ainsi, pour le Québec, 4 864 élèves ont été sélectionnés dans un premier temps, 124 élèves ont été exclus, et 390 ne se sont pas présentés le jour de l'épreuve (4 864 – 124 – 390 = 4 350).

RÉFÉRENCES

- Brown, M. (1999). Problems of interpreting international comparative data. *Oxford Studies in Comparative Education*, 9(1), 183–205.
- Martin, M.O., Mullis, I.V.S., & Chrostowski, S.J. (Éds.). (2004). *TIMSS 2003 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Repéré le 15 janvier, 2006, de <<http://timss.bc.edu/timss2003i/technicalD.html>>.
- Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Chrostowski, S.J. (2004). *Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Repéré le 15 janvier, 2006, de <<http://timss.bc.edu/timss2003i/mathD.html>>.

Jean-Guy Blais, professeur titulaire à l'Université de Montréal et directeur de l'équipe de recherche GRIÉMÉtic (<http://www.griemetic.ca/accueil.aspx>), s'intéresse aux systèmes, technologies, et modèles de mesure pour l'évaluation en éducation et formation.