



Organisation  
des Nations Unies  
pour l'éducation,  
la science et la culture

INSTITUT  
*de*  
STATISTIQUE  
*de l'UNESCO*

DOCUMENT TECHNIQUE N° 4



**MESURER LA PARTICIPATION SCOLAIRE** : Analyse de  
la qualité des données et de la méthodologie en dix études

# MESURER LA PARTICIPATION SCOLAIRE :

## Analyse de la qualité des données et de la méthodologie en dix études

---



## UNESCO

L'acte constitutif de l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO) a été adopté par 20 pays lors de la Conférence de Londres tenue en novembre 1945, et celui-ci est entré en vigueur le 4 novembre 1946. Actuellement, l'Organisation regroupe 193 États membres et sept membres associés.

Le but principal de l'UNESCO est de contribuer au maintien de la paix et de la sécurité dans le monde en resserrant, par l'éducation, la science, la culture et la communication, la collaboration entre nations, afin d'assurer le respect universel de la justice, de la loi, des droits de l'homme et des libertés fondamentales pour tous, sans distinction de race, de sexe, de langue ou de religion, que la Charte des Nations Unies reconnaît à tous les peuples.

Pour remplir son mandat, l'UNESCO assume cinq fonctions principales : 1) les études prospectives sur l'éducation, la science, la culture et la communication en vue du monde de demain ; 2) le perfectionnement, la transmission et le partage des connaissances grâce à des activités de recherche, de formation et d'enseignement ; 3) les actions normatives visant à élaborer et à adopter des instruments internes et des recommandations statutaires ; 4) l'expertise, par le biais de la coopération technique, apportée aux États membres pour leurs politiques et projets de développement et 5) l'échange d'informations spécialisées.

Le siège social de l'UNESCO est situé à Paris en France.

### Institut statistique de l'UNESCO

L'Institut de statistique de l'UNESCO (ISU), qui forme le bureau statistique de l'UNESCO, agit comme organisation dépositaire de l'ONU de statistiques à l'échelle mondiale dans les domaines de l'éducation, de la science et de la technologie, de la culture et des communications.

L'ISU a été mis sur pied en 1999. Il a été créé pour améliorer le programme statistique de l'UNESCO et pour élaborer et diffuser en temps opportun des statistiques fiables, précises et pertinentes en matière politique qu'exigent les environnements social, politique et économique d'aujourd'hui, en constante évolution et de plus en plus complexes.

L'ISU a son siège à Montréal au Canada.

Publié en 2010 par :

Institut de statistique de l'UNESCO  
B.P. 6128, succursale Centre-ville  
Montréal, Québec H3C 3J7  
Canada

Tél. : (1 514) 343-6880  
Télééc. : (1 514) 343-5740  
Courriel : [publications@uis.unesco.org](mailto:publications@uis.unesco.org)  
<http://www.uis.unesco.org>

ISBN 978-92-9189-087-3  
Réf. : UIS/TD/10-01

©UNESCO-ISU 2010

Conception de la page couverture : JCNicholls Design  
Imprimé par : OACI, Montréal

Les auteurs assument la responsabilité du choix et de la présentation des faits contenus dans ce document et les opinions qui sont exprimées aux présentes n'émanent pas nécessairement de l'UNESCO et n'engagent pas l'Organisation.

## Remerciements

Ce document technique a été préparé par Diana Maria Stukel (Westat) et Yassamin Feroz-Zada (Institut de cardiologie de Montréal) qui, au moment de la rédaction, étaient à l'emploi de l'Institut de statistique de l'UNESCO (ISU).

Les auteurs souhaitent remercier Tim Holt, professeur (émérite) de l'Université de Southampton, pour les nombreux échanges, commentaires et suggestions fructueux portant sur une version antérieure de ce document, ainsi que d'autres réviseurs, dont Thomas Buettner (Division de la population des Nations Unies), Friedrich Huebler (ISU), Mamadou Thiam (Banque mondiale) et Said O. Voffal (ISU).

L'ISU aimerait exprimer sa gratitude à Alfredo Alinga (Macro International), Friedrich Huebler, Gerhard Heilig (Division de la population des Nations Unies) et Thomas Buettner pour l'aide consentie afin d'éclaircir les enjeux relatifs aux données sous-jacentes.



## Table des matières

	Page
Remerciements .....	iii
<b>1. Introduction .....</b>	<b>7</b>
<b>2. Comparaison et décomposition globales des deux estimateurs.....</b>	<b>11</b>
Étude 1. Comparaison globale entre le TNS+ et le TNF+.....	11
Étude 2. La désagrégation du TNS+ (provenant de sources administratives et des recensements de la population) et du TNF+ (provenant des enquêtes-ménages) en éléments constituants.....	14
<b>3. Enquêtes portant sur les données démographiques de la Division de la population des Nations Unies .....</b>	<b>23</b>
Étude 3. Les effets de l'utilisation de l'interpolation de Sprague sur les données de la Division de la population des Nations Unies.....	23
Étude 4. Les répercussions des révisions des <i>Perspectives sur la population mondiale</i> sur le TNS+ .....	29
<b>4. Enquêtes relatives à la combinaison de données d'inscription provenant de sources administratives à des données sur la population issues de recensements .....</b>	<b>32</b>
Étude 5. Les répercussions sur la construction du TNS+ de l'utilisation de deux sources de données distinctes.....	32
<b>5. Enquêtes relatives aux données sur la fréquentation provenant des enquêtes-ménages .....</b>	<b>35</b>
Étude 6. De quelle façon la stratification <i>a posteriori</i> des éléments de poids des enquête-ménages se répercute-t-elle sur les estimations des totaux ? ..	35
Étude 7. Erreurs non liées à l'échantillonnage : un aperçu de la non-réponse des enquête-ménages au niveau de l'EDS .....	37
Étude 8. Erreurs d'échantillonnage : Un aperçu des erreurs-types et des coefficients de variation pour l'EDS.....	39
Étude 9. L'impact des modalités de saisie de la fréquentation de l'EDS dans divers pays.....	41
<b>6. Enquêtes relatives à la déclaration de l'âge à l'échelle des trois sources de données (enquête, administrative et recensement).....</b>	<b>44</b>
Étude 10. L'impact de la façon dont l'âge est saisi en matière de fréquentation par opposition à l'inscription .....	44
<b>7. Conclusions .....</b>	<b>50</b>
<b>Références .....</b>	<b>52</b>

## Liste des tableaux

Tableau 1.	Comparaison globale du TNS+ et du TNF+ .....	13
Tableau 2.	Définition par pays d'une année scolaire pour le niveau primaire .....	13
Tableau 3.	Comparaison des numérateurs (pour les enfants d'âge scolaire primaire)...	16
Tableau 4.	Comparaison des dénominateurs .....	17
Tableau 5.	Comparaison de la répartition de l'âge scolaire primaire telle qu'évaluée par l'EDS et la Division de la population des Nations Unies .....	20
Tableau 6.	Comparaison du nombre d'enfants d'âge scolaire primaire inscrits au nombre d'enfants qui fréquentent l'école par âge, tels qu'estimés respectivement par l'ISU et l'EDS .....	22
Tableau 7.	Les effets de l'interpolation de Sprague sur les estimations selon un âge unique des populations d'âge scolaire primaire .....	24
Tableau 8.	Les effets de l'interpolation de Sprague sur l'estimation des totaux des populations d'âge scolaire primaire (combiné) .....	27
Tableau 9.	Les effets sur le TNS+ des révisions de recensement .....	31
Tableau 10.	Les répercussions de la stratification <i>a posteriori</i> sur les estimations .....	36
Tableau 11.	Taux de réponse des ménages pour l'EDS .....	38
Tableau 12.	Taux de réponse pour l'item sur la fréquentation (fréquentation de l'école au cours de l'année scolaire actuelle/toujours à l'école) pour l'EDS .....	38
Tableau 13.	Erreur-types et coefficients de variation (CV) pour l'EDS .....	40
Tableau 14.	Comparaison entre le TNS+ et le TNF+ pour les pays où tant la question sur la fréquentation courante que celle sur la fréquentation passée ont été posées dans le cadre de l'EDS .....	41
Tableau 15.	Comparaison entre le TNS+ et le TNF+ pour les pays où seule la question sur la fréquentation courante a été posée dans le cadre de l'EDS .....	41
Tableau 16.	Comparaison des modalités de saisie de l'âge pour les données relatives à l'inscription et à la fréquentation .....	45
Tableau 17.	Définition de l'âge scolaire primaire par pays .....	45
Tableau 18.	Comparaison du TBF et du TNS afin d'enquêter sur une possible attribution fautive de l'âge par les sources de données administratives .....	46

## 1. Introduction

L'enseignement primaire universel (EPU) est un objectif commun du cadre de l'Éducation pour tous (ÉPT) et des Objectifs du millénaire pour le développement (OMD). Pour effectuer un suivi des progrès vers l'atteinte de cet objectif, il est nécessaire d'utiliser des méthodes universelles afin de mesurer la participation scolaire primaire et de compter le nombre d'enfants non scolarisés.

En se fondant sur des sources de données divergentes, l'Institut de statistique de l'UNESCO (ISU) et l'UNICEF produisent actuellement des estimations différentes de la proportion des enfants non scolarisés. Pour effectuer le suivi des taux de personnes « à l'école », l'ISU se sert de données d'inscriptions provenant de sources administratives, qui font l'objet d'une collecte chaque année auprès de tous les pays du monde, et les combine aux données des recensements de la population recueillies par la Division de la population des Nations Unies. Par contre, l'UNICEF utilise les taux de fréquentation provenant de deux enquêtes-ménages, qui sont effectuées *grosso modo* tous les cinq ans dans près de 80 pays : l'Enquête démographique et de santé (EDS ou *DHS*), commanditée par l'USAID et réalisée par Macro International, et l'Enquête en grappes à indicateurs multiples (EGIM ou *MICS*), parrainée et gérée par l'UNICEF. En outre, l'UNICEF complète les données sur la fréquentation en se servant des données d'inscriptions de l'ISU et des données issues d'autres enquêtes.

Les écarts entre les chiffres publiés par l'ISU et par l'UNICEF se sont révélés considérables, ce qui indique clairement qu'une approche unifiée s'impose. Dans le cadre d'un effort préliminaire, les deux organismes ont œuvré ensemble à la mise au point d'une estimation unique fondée sur les deux sources de données mais qui n'illustrait la situation qu'aux moments où les deux sources étaient disponibles, (UNESCO-ISU, 2005). Étant donné que la communauté des utilisateurs est intéressée par les différentes tendances au fil du temps, une estimation, fondée sur les deux sources de données qui demeure cohérente dans le temps, y compris lorsque les données des enquêtes ne sont pas disponibles, s'avère de toute évidence nécessaire. L'ISU a imparti la mise au point d'une telle méthodologie, laquelle a été livrée et passée en revue mais nécessite tout de même une validation empirique substantielle, avant de ne pouvoir être mise en œuvre pour la production d'estimations d'enfants non scolarisés (*voir Singh, 2006*).

Ce document s'attarde sur certains des enjeux soulevés par la méthodologie des enquêtes transversales. Il s'efforce, en particulier, de répondre au besoin d'une meilleure compréhension des divers aspects qualitatifs des données inhérents aux deux sources de données : les données administratives de l'ISU et les données des enquêtes-ménages utilisées par l'UNICEF. Par conséquent, les auteurs de ce rapport ont entrepris d'effectuer dix études, afin d'examiner dans quelle mesure l'écart entre les estimations dérivées de deux sources distinctes est attribuable aux questions potentielles entourant la qualité des données (plutôt qu'à des différences conceptuelles). Les résultats de ces études forment les assises de ce document. Dans l'immédiat, la question de l'élaboration d'une méthodologie visant à intégrer les deux sources de données dans le but de fournir une estimation unique des enfants non scolarisés sera mise de côté, étant donné que l'amélioration de la qualité fondamentale des données reste une étape essentielle précédant la mise au point de méthodes d'estimation améliorées.

Avant d'approfondir les détails des études comme telles, il y a lieu d'aborder rapidement la définition de la « participation scolaire », qui est au cœur des efforts déployés par l'ISU et par l'UNICEF pour effectuer le suivi des enfants non scolarisés. L'une des difficultés

relatives au suivi de la participation scolaire (ou à son complément, la participation insuffisante) réside dans le fait qu'il n'existe actuellement aucun consensus international quant à la signification exacte de ce concept. Il demeure difficile de quantifier quelque chose qui n'a pas été bien défini, et cela s'accompagne d'une variété de problèmes. En outre, il n'existe présentement aucun instrument national ou international de mesure (de nature administrative, de type enquête ou autre) qui soit expressément conçu afin de saisir et de mesurer la participation scolaire. C'est pourquoi les organismes internationaux sont astreints à l'utilisation de données produites par des instruments de mesure existants qui semblent les plus appropriés en matière de saisie du concept de participation scolaire.

Par exemple, l'ISU effectue un suivi des enfants non scolarisés en se servant du complément d'inscription (tel que mesuré par des sources administratives) et ce, malgré le fait que la participation scolaire survienne sur une base continue et que l'inscription demeure un événement unique. De plus, l'inscription saisit l'« intention de participer » plutôt que la participation en soi. Les instruments de mesure utilisés pour saisir l'inscription demeurent aussi entachés de limites. Ainsi, l'ISU demande aux pays de rapporter, pour l'inscription, des chiffres cumulatifs selon l'âge et l'année d'études. Ils se voient également sollicités à fournir une définition des divers niveaux d'enseignement (c.-à-d. le primaire, le secondaire) en indiquant l'âge d'entrée et la durée des programmes nationaux. Grâce à ces renseignements, l'ISU peut construire des taux de scolarisation pour l'école primaire.

Toutefois, nombreux sont les pays qui ne disposent pas d'un seul système d'enseignement ou d'une définition du niveau primaire qui s'applique véritablement à l'ensemble du pays. En effet, il est possible, au sein d'un pays, de retrouver des différences frappantes. Lorsque cela se produit, les pays doivent rapporter leurs données en se conformant à la définition la plus courante. L'étendue avec laquelle ce « compromis de collecte » déforme les indicateurs internationaux, tels que les taux nets de scolarisation (TNS), demeure largement inexplorée. Cependant, l'ISU n'est pas en mesure de solliciter des données qui se conforment à des définitions qui varient au sein d'un même pays, étant donné que les données administratives locales ne sont souvent pas disponibles dans une forme désagrégée.

Par contre, l'UNICEF se sert du complément qu'est la fréquentation (tel que mesurée par les enquêtes-ménages) pour effectuer un suivi des enfants non scolarisés. Cela, à plusieurs égards, se révèle plus attrayant sur le plan heuristique puisque la fréquentation se déroule de façon continue et semble intuitivement plus proche de la notion de participation. Néanmoins, les instruments d'enquête grâce auxquels on souhaite saisir la fréquentation comportent des problèmes de mesure. Par exemple, les carences relatives à la couverture de la population peuvent s'avérer un handicap pour saisir la notion de participation. Telle qu'elle est conçue, l'EDS n'englobe pas la population sans-abri et, ce faisant, elle sous-estime la population non scolarisée étant donné que seuls les enfants habitant dans les ménages et qui ne fréquentent pas l'école sont considérés comme étant non scolarisés. Dans certains pays, il se peut que cela ne concerne qu'un nombre très faible d'enfants mais, dans d'autres (en particulier dans les pays en développement), une proportion substantielle des enfants non scolarisés pourrait provenir de la population sans-abri et ne pas être couverte par les enquêtes.

Il s'agit là d'une limite importante. À cela s'ajoute le fait qu'il est possible que certains enfants vivant en institutions (p. ex. les orphelinats ou les maisons d'hébergement) présentent des niveaux plus élevés de participation scolaire que la population générale des ménages. En outre, en ce qui concerne les pays comportant une vaste population d'enfants sans-abri (en présumant que ceux-ci ne fréquentent pas l'école), il est probable

que les estimations sur la fréquentation soient surévaluées puisque, pour un numérateur, la « population fréquentant l'école » serait moins faussée que la population d'âge primaire à titre de dénominateur, étant donné que cette dernière exclut les enfants sans-abri de ses composantes. En interprétant les études qui suivent, il y a lieu de garder à l'esprit que l'utilisation des enquêtes-ménages s'accompagne de limites substantielles dans les pays où une proportion importante des enfants d'âge primaire reste sans-abri, sauf si une correction brute demeure possible.

***Recommandation 1*** : *Lors de l'utilisation des données d'enquêtes-ménages pour produire des estimations de non-participation scolaire, une marge pour les enfants sans-abri devrait être prévue partout où cette population représente un facteur d'importance.*

Le recours aux EDS soulève une autre question ayant rapport avec la mesure, en particulier en ce qui concerne les questions relatives à la « fréquentation actuelle de l'école » et la « fréquentation passée (durant une journée ou plus) au cours de l'année scolaire antérieure ». Cette dernière a pour but de saisir les étudiants qui, de façon temporaire, n'ont pas fréquenté l'école pour cause de maladie ou d'autres motifs. Mais des problèmes évidents surgissent dans la foulée de cette notion exagérée selon laquelle le simple fait de fréquenter l'école durant une seule journée au cours de la dernière année scolaire pourrait constituer une indication de participation scolaire. Et pourtant, quel pourrait être le nombre raisonnable minimal de jours de présence à l'école en vertu duquel on pourrait affirmer que, dans les faits, un enfant participe à l'enseignement? On devrait pouvoir compter sur un certain consensus international. L'une des façons permettant d'en arriver à une définition raisonnable de la participation scolaire pourrait être d'étudier les registres de fréquentation scolaire à l'échelle du pays et d'y analyser les diverses tendances quant à la fréquence et à la répartition de la fréquentation.

De la même manière, on semble constater l'existence d'un niveau très faible de compréhension des configurations de participation partielle ou de non-participation et de leur étendue à l'échelle du globe. Certains enfants peuvent être inscrits et ne jamais fréquenter l'école, d'autres peuvent la fréquenter durant un certain temps (avec ou sans une inscription officielle) ou des enfants peuvent fréquenter l'école sur une base régulière et continue. La non-fréquentation peut se révéler prévisible (p. ex. durant la saison des récoltes) ou sporadique et elle peut s'avérer intense ou sommaire (tant en ce qui concerne la durée de la non-fréquentation que le nombre de personnes visées par celle-ci). L'exactitude de la mesure de la participation scolaire dépend de ces configurations, et toute modification apportée à la cueillette des données ou aux questionnaires se doit de prendre en compte ces configurations de non-fréquentation. Conséquemment, il est essentiel non seulement de travailler à peaufiner le concept de la participation mais aussi à obtenir des données probantes empiriques sur les modalités actuelles de participation. Toutefois, il y a lieu de noter qu'on retrouve des différences considérables d'un pays à l'autre quant à la qualité des sources d'information (p. ex. des registres de fréquentation bien tenus), le fonctionnement du système d'éducation et les modèles de participation. Quoi qu'il en soit, tout renseignement, y compris les études sur une petite échelle à partir des dossiers d'un certain nombre d'écoles ou les enquêtes détaillées, peut se révéler utile.

**Recommandation 2 :** *Les organismes internationaux devraient s'entendre sur une définition de la participation / non-participation scolaire et fournir des lignes directrices et des barèmes afin de mesurer les populations non scolarisées. L'obtention de renseignements empiriques sur les modèles d'une participation partielle ou d'une non-participation contribuerait à l'élaboration du concept et des mesures de la participation scolaire. De plus, comme les instruments existants utilisés pour mesurer le nombre d'élèves inscrits et la fréquentation présentent diverses limites, les organismes internationaux devraient envisager l'utilisation d'autres sources de données. Par exemple, on pourrait se servir des registres nationaux de fréquentation (qui existent dans de nombreux pays). Il y aurait possiblement lieu de mener une enquête afin de découvrir dans quelle mesure ces renseignements seraient accessibles de façon à la fois complète, opportune et précise.*

Le reste de ce document présente les résultats des dix études associées aux différents aspects de la qualité des données issues de sources administratives, d'enquêtes-ménages et de recensements de la population. Il est fort peu probable qu'une source unique de données soit appropriée pour un éventail disparate de pays. Par conséquent, cette série d'études de qualité formule des recommandations pour le renforcement des sources de données de façon à en permettre l'utilisation la plus vaste qui soit. Si l'on veut obtenir une amélioration générale de la qualité, les recommandations se doivent d'être universelles plutôt que d'être propres à chaque pays.

Le chapitre 2 présente une comparaison générale des deux estimateurs : le taux net de scolarisation ajusté (TNS+) et le taux net de fréquentation scolaire ajusté (TNF+). Ensuite suivent une série d'études portant principalement sur des enjeux particuliers relatifs aux données démographiques de la Division de la population des Nations Unies (chapitre 3), aux données sur la fréquentation provenant de sources administratives (chapitre 4) et aux données sur la fréquentation dont les enquêtes-ménages constituent la source (chapitre 5). Les résultats d'une enquête spéciale sur l'âge indiqué dans les données apparaissent au chapitre 6, suivis par certaines conclusions au chapitre 7.

## 2. Comparaison et décomposition globales des deux estimateurs

Au départ, on a proposé que quelque 40 pays qui, chacun, comptait plus d'un million d'enfants non scolarisés d'âge primaire et / ou un TNS inférieur à 75 %, participent à l'étude. Selon le *Rapport mondial de suivi* (2002) sur l'Éducation pour tous (ÉPT), les pays choisis étaient confrontés aux défis les plus considérables quant à l'atteinte des objectifs de l'ÉPT, en ce qui a trait à l'accès et à la participation scolaire primaire. Étant donné les contraintes relatives à la disponibilité des données, l'ensemble des 40 pays a ensuite été réduit à un groupe final n'en comptant que 19. Les pays n'étaient pris en considération que si les données sur les taux de fréquentation pouvaient être obtenues soit par l'entremise de l'EDS ou de l'EGIM pour une année postérieure à 1999, et que les données administratives d'inscription correspondantes de l'ISU étaient disponibles pour cette même année.

Parmi les 19 pays, 16 disposaient de l'EDS et 3 de l'EGIM comme sources de données d'enquête. Lors de discussions avec l'UNICEF, au moment de finaliser le présent document, l'existence d'un enjeu de codage pour certaines des variables d'importance cruciale des trois pays visés par l'EGIM (Tchad, Niger et Sénégal) a été découverte, ce qui donnait l'impression que beaucoup plus de valeurs étaient manquantes qu'en réalité. Étant donné les circonstances de cette découverte, le travail empirique qui avait été réalisé pour ces trois pays n'a pu être reproduit et, malheureusement, il fut impossible de les inclure dans ce rapport. La liste finale des 16 pays couverts par l'EDS qui ont été mis à contribution dans le cadre de l'analyse apparaît au tableau 1.

### Étude 1. Comparaison globale entre le TNS+ et le TNF+

Afin de dégager des tendances générales des différences entre le taux net de scolarisation ajusté (TNS+), issu des sources administratives et des recensements, et le taux net de fréquentation scolaire ajusté (TNF+), obtenu des sources d'enquêtes-ménages, il est essentiel de comparer ces deux estimateurs.

Le TNS+ se définit comme étant :

$$\text{TNS+} = 100 \frac{\text{Nombre d'enfants d'âge scolaire primaire inscrits à l'enseignement primaire ou secondaire}}{\text{Nombre d'enfants d'âge scolaire primaire}}$$

Cela diffère quelque peu de la définition habituelle du taux net de scolarisation (TNS), lequel n'englobe au numérateur que les enfants d'âge scolaire primaire inscrits à l'école primaire (à l'exclusion des enfants d'âge scolaire primaire inscrits à l'école secondaire) (*voir UNESCO-ISU, 2007 pour plus de détails sur le TNS*). Les sources administratives nationales (c.-à-d. les ministères de l'Éducation ou les bureaux nationaux de la statistique) fournissent chaque année à l'ISU des données pour le numérateur. Quant au dénominateur, l'ISU utilise les chiffres fournis par la Division de la population des Nations Unies (DPNU), qui proviennent des données des recensements de la Division de statistique des Nations unies (DSNU). Ensuite, la Division de la population des Nations Unies produit des estimations pour les années intercensitaires pour lesquelles aucune donnée n'est disponible.

Dans le même ordre d'idées, le TNF+ se définit comme étant :

$$\text{TNF+} = 100 \frac{\text{Nombre d'enfants d'âge scolaire primaire fréquentant l'enseignement primaire ou secondaire}}{\text{Nombre d'enfants d'âge scolaire primaire}}$$

Tout comme pour le TNS+, il y a lieu de noter que cette définition diffère de la définition standard du taux net de fréquentation scolaire (TNF) en ce sens qu'elle inclut dans le numérateur les enfants d'âge scolaire primaire qui fréquentent l'école secondaire (*voir UNICEF, 2008 pour plus de détails sur le TNF.*) En ce qui concerne le TNF+, tant le numérateur que le dénominateur sont évalués à partir des enquêtes de l'EDS ou de l'EGIM, et ces estimations intègrent les poids à la fois pour le numérateur et le dénominateur.

Avant la mise au point de la méthodologie conjointe ISU/UNICEF, l'estimation de la proportion des enfants non scolarisés utilisée par l'ISU était de 100-TNS, alors que l'UNICEF utilisait 100-TNF. Ces dernières années, l'ISU a utilisé 100-TNS+. Exclure les enfants d'âge scolaire primaire inscrits ou fréquentant l'école secondaire gonfle artificiellement le nombre réel des enfants non scolarisés, ainsi il est justifié de ne plus utiliser le TNS et le TNF dans le calcul des estimations.

Quant aux enquêtes de l'EDS et de l'EGIM, les périodes de cueillette s'étendaient souvent sur deux années civiles ou plus (partielles) et, par la même occasion, sur une année scolaire à l'école primaire ou plus (*voir le **tableau 1** pour les périodes d'enquête et le **tableau 2** pour les définitions, par pays, de ce qu'est une année scolaire*). Afin de rendre le TNF+ comparable au TNS+, ce dernier a été calculé à l'aide de taux combinés fondés sur des valeurs calculées au prorata de plusieurs années scolaires correspondant le plus étroitement possible à la période de cueillette de données de l'enquête. Par exemple, au Bangladesh, l'année scolaire s'écoule de janvier à décembre. Dans ce pays, l'EDS s'est déroulée de novembre 1999 à avril 2000. Par conséquent, un taux de scolarisation combiné, utilisant à la fois les valeurs de 1999 et de 2000, a été déterminé en se servant de valeurs calculées au prorata de chaque composante, proportionnellement à la taille de l'échantillon propre aux deux derniers mois de l'enquête en 1999 (deux sixièmes) par opposition aux quatre premiers mois de l'enquête de 2000 (quatre sixièmes). On notera avec intérêt qu'une analyse distincte a été effectuée pour étudier la mesure avec laquelle les valeurs composantes de l'année à titre individuel sur laquelle étaient fondés les taux combinés de TNS+ (1999 et 2000) ne différaient pas substantiellement les unes des autres, afin de s'assurer que le taux combiné n'était pas en soi une source éventuelle de biais. Pour huit des neuf pays où on avait fait appel à la combinaison de taux, les valeurs de l'année à titre individuel se rapprochaient considérablement les unes des autres. Au Mozambique, les valeurs différaient substantiellement pour les deux années (62,82 % par rapport à 71,04 %), mais seule une très faible fraction du chiffre le plus élevé était utilisé, ce qui suscitait un taux combiné très proche du chiffre plus faible (62,91 %).

Le tableau 1 présente les résultats de la comparaison TNS+/TNF+ pour les 16 pays visés par l'EDS. Pour ce qui est du Bangladesh, la valeur absolue de la différence entre le TNF+ et le TNS+ était supérieure à 10 %. Un complément d'enquête s'impose pour mettre en lumière les causes de cette différence. En concluant qu'au Bangladesh, le TNS+ peut s'avérer un estimateur plus crédible que le TNF+, l'étude 9 fournit une explication plausible. Dans le cas de trois autres pays visés par l'EDS (Burkina Faso, Indonésie et Tanzanie), la valeur absolue des différences tournait autour de tout juste 5 %, alors que pour l'Égypte, elle se situait à environ 8 %.

**Tableau 1. Comparaison globale du TNS+ et du TNF+**

Pays	Années scolaires sur lesquelles se fonde le TNS+ combiné	TNS+ combiné	Enquête	Période d'enquête	TNF+	Différence en pourcentage entre le TNS+ et le TNF+ combinés	2 * Erreur-type (méthode jackknife) du TNF+ (telle que rapportée au tableau 13)
Bangladesh	1999, 2000	93,05	EDS	11/99-04/00	79,59	13,46	1,644
Burkina Faso	2003, 2004	38,60	EDS	06/03-12/03	32,33	6,27	3,122
Côte d'Ivoire	1998, 1999	53,67	EDS	09/98-03/99	52,15	1,52	5,396
Égypte	2000	93,50	EDS	02/00-04/00	85,50	8,00	1,134
Ghana	2003, 2004	59,80	EDS	07/03-11/03	61,22	-1,42	2,480
Guinée	1999	43,16	EDS	04/99-08/99	39,95	3,21	*
Inde	1999, 2000, 2001	82,34	EDS	11/98-07/00	82,33	0,01	0,702
Indonésie	2002, 2003	100,86	EDS	01/02-04/03	95,25	5,61	0,690
Kenya	2003	77,36	EDS	04/03-09/03	74,99	2,37	2,044
Mali	2001	42,98	EDS	01/01-06/01	39,03	3,95	2,842
Mozambique	2003, 2004	62,91	EDS	08/03-01/04	59,85	3,06	2,060
Namibie	2000	74,23	EDS	09/00-12/00	78,60	-4,37	1,860
Nigéria	2003	62,09	EDS	03/03-08/03	61,99	0,10	3,386
Rwanda	2000, 2001	71,10	EDS	06/00-12/00	71,94	-0,84	1,448
Tanzanie	1999	47,70	EDS	09/99-11/99	53,83	-6,13	3,410
Viet Nam	2002	96,06	EDS	09/02-12/02	96,27	-0,21	1,254

**Tableau 2. Définition par pays d'une année scolaire pour le niveau primaire**

Pays	Mois du début	Mois de la fin
Bangladesh	1	12
Burkina Faso	10	6
Côte d'Ivoire	10	6
Égypte	9	6
Ghana	9	7
Guinée	10	6
Inde	4	3
Indonésie	7	6
Kenya	1	12
Mali	10	6
Mozambique	1	12
Namibie	1	12
Nigéria	9	7
Rwanda	9	6
Tanzanie	1	12
Viet Nam	9	5

L'erreur-type de l'estimateur du TNF+ fondé sur l'enquête est calculé au tableau 13 en faisant appel à la technique *jackknife*; dans le tableau 1, le double de cette valeur est rapporté dans la dernière colonne. On constate que la valeur absolue de la différence en pourcentage entre le TNF+ et le TNS+ (valeur absolue de l'avant dernière colonne) est souvent supérieure (pour 9 des 15 pays) que deux erreurs-types du TNF+ (dernière colonne), ce qui souligne que les différences apparentes ne sont pas tout simplement le résultat d'erreurs d'échantillonnage de l'estimation fondée sur l'enquête mais plutôt un problème sous-jacent de biais non associé à l'échantillonnage dans l'une des sources ou dans les deux. Cela semble généralisé et ce, même pour de nombreux pays où les différences demeurent modestes.

## **Étude 2. La désagrégation du TNS+ (provenant de sources administratives et des recensements de la population) et du TNF+ (provenant des enquêtes-ménages) en éléments constituants**

Afin d'approfondir l'enquête sur les différences entre le TNS+ et le TNF+ pour les 16 pays visés par l'EDS, les valeurs de numérateur et de dénominateur ont été désagrégées.

Premièrement, il y a lieu de prendre note que l'EDS fournit des fichiers de micro-données contenant des « poids d'enquête transformés » (« *scaled survey weights* ») plutôt que les « poids originaux de base de sondage ». On obtient des poids « transformés » de l'enquête en modifiant les éléments de poids originaux de la base de sondage de façon à ce que leur somme totalise la taille globale de l'échantillon plutôt que celle de la population. Le recours aux poids « transformés » est devenue monnaie courante dans le cadre d'analyses d'ensembles de données dérivées de bases de sondage complexes, en particulier en ce qui concerne les fichiers de micro-données. L'avantage que l'on en retire réside dans le fait que lorsque les estimations fondées sur celles-ci sont saisies par les logiciels de calcul statistique standards non dotées de formules spéciales d'erreur-type appropriées aux bases de sondage complexes, une estimation d'erreur-type standard est fournie qui, à tout le moins, prend partiellement en compte les divers aspects de la base de sondage complexe. Par conséquent, en fournissant des poids « transformés », les concepteurs de l'EDS voulaient vraisemblablement éviter une utilisation naïve des poids originaux contenus dans les logiciels de calcul statistique standards, ce qui susciterait de fausses erreurs-types.

De surcroît, de l'utilisation de poids « transformés » résulte des estimateurs ponctuels qui sont identiques à ceux qui sont produits lorsque l'on recourt à des poids originaux de base de sondage, *pourvu que les estimations portent sur les ratios ou les taux*. Le problème réside dans le fait que les estimations des totaux ne peuvent être obtenues correctement en se servant des poids « transformés », malgré que cela présente parfois de l'intérêt aux yeux des analystes (p. ex. le nombre total d'enfants fréquentant l'école primaire ou secondaire). Quoi qu'il en soit, en y allant d'une demande spéciale, l'ISU a pu obtenir les poids originaux de base de sondage de Macro International (l'organisme responsable d'effectuer les enquêtes d'EDS) dans 11 des 16 pays visés par l'EDS. Ce faisant, il a été possible de réaliser une analyse distincte du numérateur et du dénominateur pour ces pays (*voir les tableaux 3, 4, 5 et 6*).

Les poids de base de sondage fournis pour les onze pays n'ont pas été ajustés en fonction de la non-réponse, et le fichier obtenu ne contenait pas suffisamment de renseignements pour être en mesure de procéder à un ajustement de la non-réponse. Dans le cas de tous les pays participants, les taux de réponse à l'enquête aux niveaux du ménage et individuel étaient élevés (>95 %). Par conséquent, il y a vraisemblablement lieu

de croire qu'avec un ajustement approprié pour tenir compte de la non-réponse, les différences d'estimation ponctuelle du TNF+ (ratio) resteraient modestes par comparaison au chiffre non ajusté. Et pourtant, il importe de noter qu'on se retrouverait vraisemblablement avec un biais de dévaluation, attribuable à la non-réponse pour les estimations des totaux, et qu'une telle sous-estimation de quelques points de pourcentage compte au moment de comparer l'estimation de la population totale fondée sur l'enquête avec celle provenant d'une autre source telle que la Division de la population des Nations Unies (tel qu'illustré au tableau 4).

**Recommandation 3 :** *Les cycles futurs des enquêtes de l'EDS et de l'EGIM devraient fournir à la fois des poids originaux de base de sondage (ajustés en fonction de la non-réponse) et des poids « transformés » aux fins de l'utilisation publique des fichiers de micro-données afin de permettre aux analystes de produire des estimations de totaux. La raison invoquée par Macro International pour justifier la non-divulgence, pour certains pays ayant participé à ces enquêtes, des poids originaux de base de sondage réside dans le fait que certains modules du questionnaire fournissaient des renseignements à l'égard desquels les estimations des totaux étaient considérées comme étant de nature trop délicate (p. ex. sur le VIH/sida). Il est difficile de comprendre cette politique étant donné qu'une estimation de la somme d'une population ne risque vraisemblablement pas plus de créer un manquement au devoir de confidentialité que ne le ferait une estimation de taux. Quoi qu'il en soit, si certains items étaient considérés d'une nature extrêmement délicate, le simple fait d'empêcher, dans le cas de ces pays, la divulgation de ces items de nature délicate aurait été une meilleure stratégie que de limiter l'utilisation pleine et entière de cet ensemble de données en empêchant que ne soit divulgués les poids originaux de base de sondage.*

Le **tableau 3** présente la différence relative en pourcentage des numérateurs (nombre des enfants inscrits par opposition au nombre des enfants qui fréquentent l'école primaire ou secondaire), alors que le **tableau 4** procure la différence relative en pourcentage des dénominateurs (nombre des enfants d'âge scolaire primaire tel qu'évalué par la Division de la population des Nations Unies via les estimations de recensement mis en opposition avec les estimations fondées sur l'EDS). On notera avec intérêt que les estimations relatives à l'inscription du tableau 3 se fondent sur des chiffres combinés tout comme dans le tableau 1.

En ce qui concerne la comparaison du nombre total d'enfants d'âge scolaire primaire inscrits à l'école par rapport à ceux qui la fréquentent (tableau no 3), la différence entre les chiffres ne surprend nullement, étant donné que l'inscription et la fréquentation demeurent conceptuellement distinctes. Bien que les deux numérateurs soient utilisés pour effectuer un suivi de la participation scolaire (son complément étant la non-participation ou la non-scolarisation), ni l'un ni l'autre n'a expressément été conçu à cette fin. Par exemple, au Bangladesh, plus ou moins 15 millions d'enfants ont été inscrits à l'école en 1999/2000 mais seuls 12,5 millions d'entre eux l'ont fréquentée. Comment pouvons-nous expliquer cette différence de 2,5 millions? Ce chiffre représente-t-il avec précision les enfants qui ont été inscrits mais qui n'ont pas fréquenté l'école? Si c'est le cas, il se peut que les deux chiffres ne soient pas entachés de biais quant à ce qu'ils entendent mesurer (respectivement l'inscription et la fréquentation), mais qu'est-ce que cela signifie au juste sur le plan de la participation? Dans cette analyse, nous nous contentons d'enquêter sur la mesure avec laquelle les estimations provenant des deux sources réussissent à effectuer un bon suivi l'une de l'autre (à la fois pour le numérateur que pour le dénominateur) et ce que cela signifie en ce qui a trait aux estimations de taux apparaissant dans le tableau 1.

D'abord, nous commençons avec une analyse distincte des tableaux 3 et 4 et, ensuite, nous les prenons en compte ensemble et examinons les résultats.

Dans le tableau 3, huit pays (Bangladesh, Côte d'Ivoire, Égypte, Indonésie, Mozambique, Namibie, Rwanda et Tanzanie) présentent des valeurs qui ont été mises en évidence là où les différences relatives en pourcentage (en chiffres absolus) sont supérieures à 10 % mais inférieures à 25 %. Dans tous les pays, à l'exception de l'Indonésie et de la Tanzanie, les valeurs sont positives, ce qui montre que les chiffres relatifs à l'inscription restent substantiellement plus élevés que ceux sur la fréquentation. En Indonésie et en Tanzanie, c'est l'inverse qui est vrai.

Il n'y a qu'un seul pays où l'écart excède 25 %, soit le Vietnam (47,2 %). Le motif derrière cette situation sera abordé de façon plus détaillée ci-dessous. Bien que des éléments de poids originaux étaient disponibles pour la Guinée, ils ont été exclus de l'analyse pour ce tableau (et de l'analyse des tableaux 6, 10, 13, 17 et 18) à cause d'une erreur de calcul qui a été découverte juste avant la publication.

**Tableau 3. Comparaison des numérateurs (pour les enfants d'âge scolaire primaire)**

Pays	Nombre total d'inscriptions (chiffre de source administrative de l'ISU fondé sur un chiffre combiné)	Nombre total au niveau de la fréquentation (estimation de l'EDS)	Différence en pourcentage (par rapport aux élèves inscrits)	Valeur absolue de la différence entre A et B	2 * Erreur-type de l'estimation de l'EDS	D<N?
	A	B	C	D	E	F
Bangladesh	15 020 499	12 467 164	-17,00	2 553 334	493 828	FAUX
Côte d'Ivoire	1 473 852	1 313 722	-10,86	160 130	202 888	VRAI
Égypte	7 340 000	6 530 655	-11,03	809 345	313 856	FAUX
Indonésie	25 184 874	29 526 814	17,24	4 341 939	1 630 718	VRAI
Mozambique	2 317 869	1 842 006	-20,53	475 863	98 438	FAUX
Namibie	283 802	225 569	-20,52	58 232	26 404	FAUX
Nigéria	13 211 302	12 030 214	-8,94	1 181 088	1 247 614	VRAI
Rwanda	1 046 634	910 074	-13,05	136 559	35 414	FAUX
Tanzanie	3 105 435	3 443 945	10,90	338 510	386 822	VRAI
Viet Nam	8 498 039	4 487 110	-47,20	4 010 929	362 864	FAUX

**Tableau 4. Comparaison des dénominateurs**

Pays	Nombre total d'enfants d'âge scolaire primaire (chiffre du recensement de la DPNU)	Nombre total d'enfants d'âge scolaire primaire (estimation de l'EDS)	Différence relative en pourcentage (par rapport à la DPNU)	Valeur absolue de la différence entre A et B	2 * Erreur-type de l'estimation de l'EDS	D<N?
	A	B	C	D	E	F
Bangladesh	16 141 818	15 671 789	-2,91	470 028	596 678	VRAI
Côte d'Ivoire	2 746 155	2 522 098	-8,16	224 056	287 116	VRAI
Égypte	7 850 375	7 638 398	-2,70	211 976	363 804	VRAI
Guinée	1 312 388	1 390 773	5,97	78 385	59 226	FAUX
Indonésie	24 970 629	30 999 584	24,14	6 028 955	1 691 758	FAUX
Mozambique	3 684 618	3 082 436	-16,34	602 181	164 236	FAUX
Namibie	382 332	286 563	-25,05	95 768	32 236	FAUX
Nigéria	21 276 574	19 549 112	-8,12	1 727 462	1 625 202	FAUX
Rwanda	1 471 994	1 264 968	-14,06	207 026	42 588	FAUX
Tanzanie	6 510 555	6 477 654	-0,51	32 900	619 628	VRAI
Viet Nam	8 846 251	4 662 966	-47,29	4 183 285	388 504	FAUX

Le tableau 4 compare les dénominateurs. On s'attend à près d'une correspondance entre les chiffres de la Division de la population des Nations Unies et ceux de l'EDS (à tout le moins sur le plan conceptuel) puisqu'ils visent tous deux à évaluer le nombre d'enfants d'âge scolaire primaire. Toutefois, en ce qui concerne le Vietnam, la différence relative se révèle extrêmement élevée (47 %). Cela est vraisemblablement attribuable à un problème associé aux poids d'échantillonnage de l'EDS : dans les tableaux 3 et 4, on retrouve trois sources indépendantes de données (de source administrative de l'ISU, du recensement de la Division de la population des Nations Unies et de l'EDS) mais un facteur commun qui est l'EDS ; il est donc vraisemblable de croire que le problème découle de cette source de données. Étant donné que les deux tableaux 3 et 4 présentent des valeurs identiques (47 %) pour le Vietnam, il semblerait alors que le Vietnam possède des poids qui sont décentrés de l'ordre d'un facteur de 2; c'est-à-dire que les poids devraient être doublés si l'on veut qu'ils représentent correctement la population. Lors de consultations avec Macro International, nous avons été informés que l'enquête EDS de 2002 au Vietnam avait été fondée sur l'EDS antérieure de 1997, laquelle en retour s'apparentait à un sous-échantillon du *Multi-Round Demographic Survey* de 1996. Quoi qu'il en soit, Macro International a décidé, pour l'EDS de 2002, de ne pas majorer les poids finaux de l'inverse du taux de sous-échantillonnage étant donné que leur principal intérêt consistait à produire des estimations de ratio et que le fait d'omettre cette étape n'y changerait rien. Afin d'illustrer ce point, disons que bien que les écarts entre les composantes du numérateur et du dénominateur du TNF+ et du TNS+ pour le Vietnam soient élevés, les ratios réels demeurent très près l'un de l'autre avec une différence de moins de 1 % (voir tableau 1).

Si les poids avaient été majorés de l'inverse des taux vietnamiens de sous-échantillonnage, cela aurait généré des estimations des totaux fondées sur les enquêtes à la fois dans les tableaux 3 et 4, cela nous aurait rapproché de l'estimation issue de la source de rechange correspondante (chiffre sur la fréquentation du tableau 3 et nombre d'enfants d'âge scolaire primaire dans le tableau 4).

**Recommandation 4 :** *Si les poids originaux de base de sondage étaient ajoutés aux fichiers de micro-données (aux fins de l'utilisation publique), tel que cela est suggéré dans la recommandation 3, il faudrait alors que cela se reflète sur tout sous-échantillonnage qui serait prélevé entre les cycles de l'EDS (comme dans le cas du Vietnam en 2002) en majorant les poids de l'inverse du taux de sous-échantillonnage. Si cela n'est pas fait, les estimations des totaux pourraient être entachées d'imprécisions considérables. Même si les poids originaux de base de sondage ne sont pas mis à la disposition du fichier de micro-données public, cela devrait être fait étant donné que les taux de sous-échantillonnage pourraient se révéler différents à une échelle sous-nationale et que cela pourrait se répercuter sur les estimations des ratios.*

Le tableau 4 illustre quatre pays (Indonésie, Mozambique, Namibie et Rwanda) qui affichent des différences en valeurs absolues qui varient de 10% à 25 %. On peut aussi observer dans le tableau 3 des écarts prononcés pour ces quatre mêmes pays. Comme pour ces pays, les estimations fondées sur les enquêtes restent à la fois communes au numérateur et au dénominateur, on pourrait alors en venir à la conclusion que les écarts sont attribuables à ces estimateurs. Cet enjeu sera abordé de façon plus approfondie dans l'étude 6 et dans les tableaux 5 et 6.

Les données de l'enquête fournissent des estimations relativement précises du TNF+ (c.-à-d. accompagnés d'erreurs-types faibles - voir tableau 13) mais elles ont tendance à être beaucoup moins précises quant à l'estimation des totaux pour les composantes du numérateur et du dénominateur des tableaux 3 et 4 (c.-à-d. ceux qui contiennent les erreurs-types plus élevées). Le ratio des deux estimations fondées sur l'enquête possède de meilleures propriétés statistiques que l'un ou l'autre des éléments constitutifs à cause de la corrélation entre le numérateur et le dénominateur. Conséquemment, le problème ne consiste pas à savoir si les numérateurs ou les dénominateurs du TNF+ font l'objet d'une estimation qui soit de bonne ou mauvaise qualité, puisque le ratio sera doté de bonnes propriétés statistiques. Il faudrait plutôt se poser la question de savoir si les numérateurs et les dénominateurs du TNF+, peu importe leur qualité, jettent une lumière quelconque sur les comparaisons correspondantes avec les valeurs du TNS+.

Dans les deux tableaux 3 et 4, on effectue une comparaison entre les différences absolues des estimations (colonne D) et le double de l'erreur-type de l'estimation de l'enquête (colonne E). Étant donné que les estimations de l'enquête et de la Division de la population des Nations Unies du tableau 4 suivent une même quantité, il est possible d'identifier des cas où l'estimation de l'enquête semble « sur la bonne voie ». La colonne F démontre que l'estimation de l'enquête retrace bien l'estimation de la Division de la population des Nations Unies dans quatre pays (Bangladesh, Côte d'Ivoire, Égypte et Tanzanie). Si l'on jette un œil sur les quatre mêmes pays dans le tableau 3, on observe que les différences en pourcentage (colonne C) sont légèrement supérieures, ou dans certains cas considérablement plus élevées, que dans le tableau 4. Par conséquent, les numérateurs semblent davantage poser problème que les dénominateurs. En conclusion provisoire, on peut avancer que même dans le cas des pays où, en ce qui concerne le dénominateur, les estimations démographiques de l'enquête s'apparentent à celles de la Division de la population des Nations Unies, il existe certaines données probantes à l'effet que, en ce qui a trait au numérateur, les données d'inscription semblent légèrement plus élevées que celles de fréquentation (et beaucoup plus importantes dans certains cas). En tout état de cause, il se peut que les données d'inscription donnent lieu à des problèmes en Côte d'Ivoire, en Égypte et en Tanzanie (la problème avec le Bangladesh ayant été résolu plus tôt) étant donné que les estimations fondées sur l'enquête pour le

dénominateur semblent crédibles et que, conséquemment, il y a lieu de croire qu'elles seront tout autant crédibles pour le numérateur.

Dans le tableau 4 (colonne C), quatre pays (Indonésie, Mozambique, Namibie et Rwanda) affichent des écarts importants entre l'estimation de la population selon l'enquête et celle de la Division de la population des Nations Unies (la question relative au Vietnam ayant été résolue plus tôt). Pour approfondir l'analyse à cet égard, dans le **tableau 5**, les données du tableau 4 ont été ventilées en fonction d'âges uniques, afin de démontrer que les écarts entre l'EDS et la Division de la population des Nations Unies pouvaient découler de certains âges posant problèmes. Dans le tableau 5, les chiffres qui révèlent une différence relative supérieure à 10 % en valeur absolue sont mis en évidence. Au Mozambique, la différence de 16 % du tableau 4 est en grande partie justifiée par des enfants âgés de 7, 9, 10 et 11 ans, bien que tous les groupes d'âges posent relativement problème. Au Rwanda, la différence est dominée par des composantes issues des enfants âgés de 7, 8, 9 et 11 ans. Par contre, en Indonésie et en Namibie, ce sont les données pour presque tous les âges qui semblent poser problème. À l'étude 6, nous explorons une solution susceptible de composer avec cet écart démographique en fonction de l'âge.

Il y a lieu de noter que chaque pays présente un âge au début de l'enseignement primaire et une durée de celui-ci qui diffèrent, selon ce qui est défini par la Classification internationale type de l'éducation (CITE) (*voir tableau 17*). Au Bangladesh, par exemple, on établit l'âge scolaire primaire à 6 à 10 ans, alors qu'en Tanzanie, il est fixé à 7 à 13 ans. Dans le tableau 5, les portées d'âges mentionnées se veulent conformes à la définition nationale de l'âge scolaire primaire et, conséquemment, diffèrent pour chaque pays.

**Tableau 5. Comparaison de la répartition de l'âge scolaire primaire telle qu'évaluée par l'EDS et la Division de la population des Nations Unies**

Pays	Âge	Nombre total d'enfants d'âge scolaire primaire (chiffre du recensement de la DPNU)	Nombre total d'enfants d'âge scolaire primaire (estimation de l'EDS)	Différence relative en pourcentage (par rapport à la DPNU)
Bangladesh (1999-2000)	6	3 334 029	2 814 202	-15,6
	7	3 288 537	3 366 785	2,4
	8	3 235 003	3 129 013	-3,3
	9	3 175 721	2 738 539	-13,8
	10	3 108 528	3 623 251	16,6
Côte d'Ivoire (1998-1999)	6	476 930	472 725	-0,9
	7	468 646	405 808	-13,4
	8	460 945	438 656	-4,8
	9	453 556	381 788	-15,8
	10	446 455	475 167	6,4
Égypte (2000)	6	1 540 452	1 541 170	0,0
	7	1 549 989	1 263 143	-18,5
	8	1 565 857	1 405 853	-10,2
	9	1 585 313	1 648 261	4,0
	10	1 608 764	1 779 971	10,6
Guinée (1999)	7	241 080	289 759	20,2
	8	231 923	242 976	4,8
	9	222 924	209 458	-6,0
	10	214 041	270 279	26,3
	11	205 230	153 673	-25,1
Indonésie (2002-2003)	7	4 135 876	5 233 917	26,5
	8	4 140 363	5 065 541	22,3
	9	4 150 236	5 048 195	21,6
	10	4 164 050	5 285 793	26,9
	11	4 180 357	4 918 402	17,7
Mozambique (2003-2004)	12	4 199 747	5 447 736	29,7
	6	569 625	518 388	-9,0
	7	554 210	477 398	-13,9
	8	539 480	494 896	-8,3
	9	525 284	396 095	-24,6
Namibie (2000)	10	511 722	443 627	-13,3
	11	498 902	314 347	-37,0
	12	485 396	437 681	-9,8
	6	60 183	44 777	-25,6
	7	58 897	39 886	-32,3
Nigéria (2003)	8	57 192	43 388	-24,1
	9	55 167	39 627	-28,2
	10	52 843	45 104	-14,6
	11	50 244	34 126	-32,1
	12	47 806	39 650	-17,1
Nigéria (2003)	6	3 774 781	3 596 905	-4,7
	7	3 673 137	3 673 261	-0,0
	8	3 580 089	3 630 518	1,4
	9	3 493 793	2 807 832	-19,6
	10	3 414 063	3 560 902	4,3
	11	3 340 711	2 279 695	-31,8

Pays	Âge	Nombre total d'enfants d'âge scolaire primaire (chiffre du recensement de la DPNU)	Nombre total d'enfants d'âge scolaire primaire (estimation de l'EDS)	Différence relative en pourcentage (par rapport à la DPNU)
Rwanda (2000)	7	240 582	202 526	-15,8
	8	241 668	208 473	-13,7
	9	243 399	180 457	-25,9
	10	246 133	233 757	-5,0
	11	250 234	188 092	-24,8
	12	249 978	251 659	0,7
Tanzanie (1999)	7	990 390	1 014 001	2,4
	8	969 267	1 029 554	6,2
	9	949 085	966 626	1,8
	10	930 143	862 876	-7,2
	11	912 748	778 281	-14,7
	12	892 293	925 771	3,8
	13	866 629	900 543	3,9
Viet Nam (2002)	6	1 681 211	771 614	-54,1
	7	1 726 794	893 587	-48,3
	8	1 771 494	963 124	-45,6
	9	1 813 072	997 762	-45,0
	10	1 853 680	1 089 947	-41,2

Les totaux de population dans le tableau 4 pour le Bangladesh, la Côte d'Ivoire, l'Égypte et la Tanzanie restent relativement semblables, mais leurs estimations de la participation dans le tableau 3 diffèrent. Par conséquent, il est digne d'intérêt d'examiner les estimations de participation et de population, en fonction de l'âge unique, qui sont fournies au **tableau 6**. Les données affichant des différences relatives supérieures à 10 % en valeur absolue sont mises en évidence. Une récurrence des problèmes rencontrés au tableau 3 apparaît pour la Côte d'Ivoire, l'Égypte et la Tanzanie et ce, pour plusieurs âges.

Étant donné qu'ils ont été mentionnés précédemment, il ne sera pas question ici des problèmes relatifs au Bangladesh et au Vietnam.

Des écarts importants ont été constatés plus tôt pour l'Indonésie et la Namibie, à la fois pour les estimations du numérateur et celles du dénominateur. Dans le tableau 6, presque tous les groupes d'âges dans ces pays présentent d'importants écarts. Dans le cas du Mozambique et du Rwanda, il n'a pas été possible d'effectuer un découpage en fonction d'âges uniques donnés.

La conclusion globale que nous tirons de l'étude 2 est la suivante : en ce qui concerne les pays où les estimations de la population d'âge scolaire primaire de l'enquête restent relativement conformes aux estimations fondées sur le recensement de la Division de la population des Nations Unies quant aux dénominateurs (Bangladesh, Côte d'Ivoire, Égypte et Tanzanie), il est probant que les données d'inscription des numérateurs sont légèrement gonflées lorsqu'on les compare aux estimations de la fréquentation (sauf pour la Tanzanie), ce qui laisserait supposer l'existence d'un problème en ce qui a trait aux données d'inscription dans ces pays. Les pays qui présentent des estimations de la population d'âge scolaire primaire qui ne sont pas cohérentes à l'échelle des deux sources de données (Indonésie, Mozambique, Namibie et Rwanda) font l'objet, plus tard dans ce document, d'une réévaluation.

**Tableau 6. Comparaison du nombre d'enfants d'âge scolaire primaire inscrits au nombre d'enfants qui fréquentent l'école par âge, tels qu'estimés respectivement par l'ISU et l'EDS**

Pays	Âge	Nombre total d'enfants inscrits (chiffre de source administrative de l'ISU fondé sur un chiffre combiné)	Nombre total d'enfants fréquentant l'école (estimation de l'EDS)	Différence relative en pourcentage (par rapports aux élèves inscrits)
Bangladesh (1999-2000)	6	3 224 841	1 886 921	-41,5
	7	3 161 639	2 624 373	-17,0
	8	3 128 934	2 612 193	-16,5
	9	2 795 148	2 354 709	-15,8
	10	2 517 674	2 988 968	18,7
Côte d'Ivoire (1998-1999)	6	235 547	142 488	-39,5
	7	262 165	192 786	-26,5
	8	268 773	262 687	-2,3
	9	252 798	230 769	-8,7
	10	226 445	265 072	17,1
Égypte (2000)	6	1 389 717	873 638	-37,1
	7	1 435 446	1 171 219	-18,4
	8	1 523 739	1 316 824	-13,6
	9	1 590 642	1 540 590	-3,1
	10	1 400 456	1 628 383	16,3
Indonésie (2002-2003)	7	4 238 414	4 804 825	13,4
	8	4 541 099	4 899 029	7,9
	9	4 436 898	4 928 233	11,1
	10	4 338 569	5 135 989	18,4
	11	4 250 128	4 755 739	11,9
Namibie (2000)	6	19 247	11 542	-40,0
	7	40 503	31 312	-22,7
	8	46 080	37 401	-18,8
	9	46 084	36 292	-21,2
	10	49 378	40 776	-17,4
Nigéria (2003)	6	1 658 828	1 312 456	-20,9
	7	2 014 657	1 983 363	-1,6
	8	2 212 151	2 386 253	7,9
	9	2 354 664	2 041 766	-13,3
	10	2 197 043	2 437 929	11,0
Tanzanie (1999)	6	2 773 959	1 868 448	-32,6
	7	145 636	218 459	50,0
	8	315 083	399 400	26,8
	9	443 522	439 165	-1,0
	10	538 700	498 130	-7,5
Viet Nam (2002)	6	564 750	556 280	-1,5
	7	558 051	694 798	24,5
	8	539 693	637 711	18,2
	9	1 539 876	688 859	-55,3
	10	1 630 674	872 795	-46,5
	8	1 780 672	935 560	-47,5
	9	1 752 685	979 612	-44,1
	10	1 793 994	1 063 458	-40,7

### **3. Enquêtes portant sur les données démographiques de la Division de la population des Nations Unies**

#### **Étude 3. Les effets de l'utilisation de l'interpolation de Sprague sur les données de la Division de la population des Nations Unies**

Pour le dénominateur de l'indicateur de TNS+, l'ISU obtient de la Division de la population des Nations Unies des données agrégées de population par groupes d'âges concernés par l'école primaire dans le pays dont il est question. Normalement, la Division de la population des Nations Unies a accès à des données par âge unique pour environ un tiers des pays du monde, alors que les deux-tiers restants sont couverts par des données présentées en fonction de groupes traditionnels de cinq ans d'âge (0 à 4 ans, 5 à 9 ans, 10 à 14 ans, etc.).

Ces données sont issues des recensements nationaux recueillis par la Division de statistique des Nations Unies (DSNU). Étant donné que la Division de la population des Nations Unies adopte une méthodologie qui est commune à tous les pays, les données reçues pour un âge unique sont regroupées en groupes traditionnels de cinq ans d'âge conformément aux données reçues des deux-tiers des pays. Ensuite, on y applique des méthodes d'estimation en faisant appel à des renseignements externes portant sur les taux de fécondité, de mortalité et de migration pour les années intercensitaires pour lesquelles aucune estimation n'existe.

Toutefois, l'ISU nécessite que ces chiffres soient présentés non pas en fonction des groupes standards de cinq ans d'âge mais plutôt en regroupement d'âges conformes à la définition de l'âge scolaire primaire propre à chaque pays. Par conséquent, les données qui sont traditionnellement diffusées par la Division de la population des Nations Unies ne peuvent être utilisées pour le calcul du TNS+. Une entente bilatérale existe en vertu de laquelle la Division de la population des Nations Unies ajoute une étape supplémentaire à ses estimations finales afin de composer avec les besoins spéciaux de l'ISU. C'est-à-dire qu'elle utilise un algorithme d'interpolation spécial, qui intègre une formule osculatrice par cinquième différence que l'on doit à Sprague, 1881, afin de ventiler les estimations finales provenant des groupes d'âge quinquennaux pour les transformer en estimations en fonction d'un âge unique. L'ISU peut ensuite regrouper les chiffres en fonction des groupements nécessaires à la production des estimations du TNS+. Ces estimations de populations par année d'âge sont aussi publiquement accessibles pour 196 pays et territoires (Division de la population des Nations Unies, 2008).

À cause de cette étape supplémentaire au niveau de l'application de l'interpolation de Sprague, une variation en sus se trouve intégrée aux estimations démographiques utilisées par l'ISU au niveau du dénominateur du TNS+. Cette étude entreprend d'enquêter sur la mesure avec laquelle cela constitue une source de préoccupation et contribue aux différences qui distinguent les estimations de l'ISU et celles fondées sur l'enquête des dénominateurs du TNS+ et du TNF+ respectivement.

On notera avec intérêt que l'analyse n'a été possible que pour un ensemble spécial de pays, souscrivant à une initiative conjointe de l'ISU/OCDE appelée programme sur les Indicateurs de l'éducation dans le monde (IEM). Le programme englobe 19 pays à revenu intermédiaire (et généralement les plus peuplés) auprès desquels l'OCDE et l'ISU ont recueilli directement des données d'âge unique, c'est-à-dire que ces données n'ont pas été obtenues de la Division de la population des Nations Unies. Sur ces 19 pays, seuls

deux (l'Égypte et l'Indonésie) font aussi partie des 16 pays sur lesquels ce document est axé. Cette étude particulière imite le processus utilisé par la Division de la population des Nations Unies pour les pays visés par l'IEM en regroupant les données en groupements standards de cinq ans d'âge et, ensuite, en appliquant l'algorithme d'interpolation de Sprague pour vérifier à quel point les données d'âge unique (deuxième colonne du **tableau 7**) qui en découlent correspondent étroitement aux estimations originales de l'IEM (troisième colonne du tableau 7). La différence relative en pourcentage apparaît dans la quatrième colonne. Les chiffres supérieurs à 10 % sont mis en évidence en gris foncé, alors que ceux qui se situent entre 5 % et 10 % sont mis en surbrillance en gris pâle.

**Tableau 7. Les effets de l'interpolation de Sprague sur les estimations selon un âge unique des populations d'âge scolaire primaire**

Âges scolaires primaires au sein des pays	Estimations démographiques, interpolation de Sprague	Estimations démographiques, IEM (données originales)	Différence relative en pourcentage de l'interpolation de Sprague (par rapport aux données originales de l'IEM)
<b>Argentine, 2003</b>			
6	670 227	687 086	-2,45
7	700 874	684 015	2,46
8	714 684	680 966	4,95
9	715 039	677 948	5,47
10	702 068	675 093	4
11	675 903	672 531	0,5
<b>Brésil 2003</b>			
7	3 551 001	3 636 128	-2,34
8	3 654 123	3 620 271	0,94
9	3 689 362	3 590 516	2,75
10	3 657 055	3 561 225	2,69
<b>Chili 2003</b>			
6	281 962	288 738	-2,35
7	296 843	290 067	2,34
8	304 727	291 175	4,65
9	306 869	291 962	5,11
10	303 330	292 488	3,71
11	294 170	292 815	0,46
<b>Chine 2002</b>			
7	18 232 357	17 914 756	1,77
8	20 304 834	18 752 106	8,28
9	22 123 665	20 082 026	10,17
10	23 803 276	26 210 044	-9,18
11	25 458 094	25 137 678	1,27
<b>Égypte 2003</b>			
6	1 348 216	1 457 444	-7,49
7	1 421 328	1 374 873	3,38
8	1 465 463	1 317 026	11,27
9	1 486 589	1 294 569	14,83
10	1 482 575	1 329 439	11,52
<b>Féd. de Russie 2003</b>			
7	1 395 548	1 403 952	-0,6
8	1 550 111	1 447 971	7,05
9	1 697 062	1 473 846	15,15

Âges scolaires primaires au sein des pays	Estimations démographiques, interpolation de Sprague	Estimations démographiques, IEM (données originales)	Différence relative en pourcentage de l'interpolation de Sprague (par rapport aux données originales de l'IEM)
<b>Inde 2004</b>			
6	24 979 127	26 412 941	-5,43
7	26 317 098	23 045 028	14,2
8	27 001 328	31 031 074	-12,99
9	27 128 710	20 868 238	30
10	26 755 786	33 855 378	-20,97
<b>Indonésie 2003</b>			
7	4 316 952	4 221 700	2,26
8	4 438 881	4 248 200	4,49
9	4 479 827	4 269 900	4,92
10	4 440 173	4 287 700	3,56
11	4 320 303	4 301 300	0,44
12	4 231 518	4 307 500	-1,76
<b>Jamaïque 2003</b>			
6	56 169	56 823	-1,15
7	59 086	56 494	4,59
8	60 476	59 735	1,24
9	60 609	59 310	2,19
10	59 530	58 010	2,62
11	57 287	55 250	3,69
<b>Jordanie, 2003</b>			
6	125 604	125 721	-0,09
7	131 030	124 461	5,28
8	133 172	133 817	-0,48
9	132 658	131 298	1,04
10	129 565	131 544	-1,5
11	123 970	124 707	-0,59
<b>Malaisie 2003</b>			
6	542 338	555 700	-2,4
7	564 049	548 500	2,83
8	571 751	543 100	5,28
9	568 287	538 300	5,57
10	553 476	531 300	4,17
11	527 138	526 200	0,18
<b>Paraguay 2003</b>			
6	126 892	129 127	-1,73
7	134 424	131 747	2,03
8	138 791	133 880	3,67
9	140 540	135 402	3,79
10	139 660	136 228	2,52
11	136 147	136 281	-0,1
<b>Pérou 2003</b>			
6	594 479	609 018	-2,39
7	622 368	607 829	2,39
8	634 963	605 885	4,8
9	635 138	603 153	5,3
10	623 106	599 843	3,88
11	599 073	596 166	0,49

Âges scolaires primaires au sein des pays	Estimations démographiques, interpolation de Sprague	Estimations démographiques, IEM (données originales)	Différence relative en pourcentage de l'interpolation de Sprague (par rapport aux données originales de l'IEM)
<b>Philippines 2003</b>			
6	2 007 719	2 022 815	-0,75
7	2 083 165	2 064 351	0,91
8	2 103 902	1 993 364	5,55
9	2 081 351	2 017 126	3,18
10	2 012 074	2 052 869	-1,99
11	1 892 630	1 848 693	2,38
<b>Sri Lanka 2003</b>			
5	271 821	302 422	-10,12
6	297 298	304 584	-2,39
7	314 745	307 459	2,37
8	325 505	310 933	4,69
9	330 921	314 892	5,09
<b>Thaïlande 2003</b>			
6	943 256	960 276	-1,77
7	988 693	960 151	2,97
8	1 011 494	965 974	4,71
9	1 016 236	973 765	4,36
10	1 003 878	979 490	2,49
11	975 380	980 118	-0,48
<b>Tunisie, 2004</b>			
6	174 858	180 860	-3,32
7	180 778	180 860	-0,05
8	186 780	180 860	3,27
9	192 703	180 860	6,55
10	198 485	207 660	-4,42
11	204 063	207 660	-1,73
<b>Uruguay 2003</b>			
6	54 988	56 346	-2,41
7	57 531	56 174	2,42
8	58 672	55 956	4,85
9	58 681	55 694	5,36
10	57 592	55 419	3,92
11	55 436	55 165	0,49
<b>Zimbabwe 2002</b>			
6	314 464	332 470	-5,42
7	326 149	315 889	3,25
8	330 751	302 544	9,32
9	330 025	294 592	12,03
10	322 931	295 864	9,15
11	308 428	315 334	-2,19
12	302 260	316 210	-4,41

En général, il semble que le processus d'interpolation de Sprague ne provoque pas de variation excessive. Néanmoins, des problèmes sont décelés à l'échelle de la plupart des groupes d'âges pour la Chine, l'Égypte, l'Inde, la Russie et le Zimbabwe, alors que seuls certains groupes d'âges posent problème dans d'autres pays (p. ex. à l'âge de cinq ans au Sri Lanka). Trois des pays du premier groupe font partie des plus peuplés de la planète (Chine, Inde et Russie) et, conséquemment, les écarts constatés pour ces pays pourraient donner lieu à des répercussions substantielles non seulement sur les estimations au niveau des pays du TNS+, mais également sur les estimations régionales du TNS+ au sein desquelles leur apport joue un rôle dominant. Étant donné que 5 de ces 19 pays exerçant une forte influence présentent une différence relative considérable accompagnée d'une forte variation supplémentaire, laquelle est entièrement attribuable à l'interpolation de Sprague, l'ISU devrait peut-être envisager d'autres modes de production d'estimations démographiques en fonction d'un âge unique.

On notera avec intérêt que bien que les répercussions de l'interpolation de Sprague sur les estimations en fonction d'un âge unique restent considérables dans certains cas, le TNS+ est généralement calculé en fonction du groupe d'âges combiné qui englobe l'âge scolaire primaire et, par conséquent, l'effet de l'interpolation de Sprague demeurera vraisemblablement plus modéré pour le groupe combiné que pour les âges uniques. Afin d'illustrer ce phénomène, une analyse (semblable au tableau 7) a été réalisée sur les groupes d'âges cumulatifs et les résultats sont présentés au **tableau 8**. On s'aperçoit que bien que les répercussions de l'interpolation de Sprague soit atténuée pour la plupart des pays par rapport au tableau 7, les effets demeurent quand même considérables pour l'Égypte et la Fédération de Russie, légèrement supérieurs à 5 %. Conséquemment, l'ISU aurait intérêt à remettre en question la notion d'utilisation de la méthode d'interpolation de Sprague.

**Tableau 8. Les effets de l'interpolation de Sprague sur l'estimation des totaux des populations d'âge scolaire primaire (combiné)**

Pays	Estimations démographiques, interpolation de Sprague	Estimations démographiques, IEM (données originales)	Différence relative en pourcentage Interpolation de Sprague (par rapport aux données originales de l'IEM)
Argentine, 2003	4 178 795	4 077 639	2,48
Brésil, 2003	17 915 795	17 648 719	1,51
Chili, 2003	1 787 901	1 747 244	2,33
Chine, 2002	109 922 226	108 096 610	1,69
Égypte, 2003	7 204 171	6 773 351	6,36
Féd. de Russie, 2003	4 642 721	4 325 769	7,33
Inde, 2004	132 182 049	135 212 659	-2,24
Indonésie, 2003	26 227 654	25 636 300	2,31
Jamaïque, 2003	353 157	345 622	2,18
Jordanie, 2003	775 999	771 548	0,58
Malaisie, 2003	3 327 039	3 243 100	2,59
Paraguay, 2003	816 454	802 665	1,72
Pérou, 2003	3 709 127	3 621 894	2,41
Philippines, 2003	12 180 841	11 999 218	1,51
Sri Lanka, 2003	1 540 290	1 540 290	0,00
Thaïlande, 2003	5 938 937	5 819 774	2,05
Tunisie, 2004	1 137 667	1 138 760	-0,10
Uruguay, 2003	342 900	334 754	2,43
Zimbabwe, 2002	2 235 008	2 172 903	2,86

Dans le passé, l'ISU a envisagé la possibilité de recueillir, directement auprès des pays par l'entremise de son enquête annuelle sur l'éducation, des données intercensitaires sur la population produites à l'échelle nationale pour les estimations en fonction d'un âge unique, éliminant ainsi la nécessité des estimations de la Division de la population des Nations Unies et de la technique d'interpolation de Sprague. Cependant, ce faisant, l'ISU se verrait accorder un rôle intenable d'arbitre, devant ainsi prendre des décisions quant à ce qui constitue des estimations démographiques provenant d'un pays qui soient acceptables, un domaine qui, de toute évidence, ne relève pas du mandat de l'ISU. En réalité, à l'occasion, les pays eux-mêmes n'arrivent pas à s'entendre ni à l'interne, à l'échelle des divers ministères, ni avec la Division de la population des Nations Unies, lorsqu'il est question de déterminer quels sont les chiffres qui sont officiels.

Plutôt que d'utiliser les groupes quinquennaux traditionnels pour tous les pays, la Division de la population des Nations Unies devrait envisager la production d'estimations pour des données d'âge unique et ce, même pour les pays qui, *a priori*, ne fournissent pas ce genre de données. Ce changement de stratégie présenterait le double avantage de prévenir la perte de renseignements (à cause du regroupement) pour les pays qui fournissent dans les faits des données à ce niveau et d'aider l'ISU à répondre à ses besoins de données particuliers.

***Recommandation 5 :*** *L'ISU devrait investir des ressources afin de travailler étroitement avec la Division de la population des Nations Unies à la formulation d'une stratégie de rechange pouvant remplacer le recours à l'interpolation de Sprague. Étant donné l'importance cruciale des estimations démographiques pour les indicateurs de l'ISU, il y aurait lieu de réfléchir sérieusement au financement d'un membre du personnel qui se consacrerait à cette question.*

Dans l'ensemble, l'un des plus importants obstacles à l'obtention en quantité suffisante de données de haute qualité sur la population réside dans le fait que de nombreux pays n'effectuent pas assez souvent de recensements de la population. Il est recommandé que les pays en réalise un au moins à tous les dix ans. Lors de la dernière édition de recensement en 2000, plus de 25 pays n'ont effectué aucun recensement; 17 de ceux-ci étaient situés en Afrique subsaharienne. De surcroît, même lorsque des recensements étaient réalisés, la qualité des données produites laissait souvent à désirer.

Parmi les problèmes-types qui empoisonnent de telles bases de données, on retrouve une sous-couverture considérable (soit pour des régions géographiques données ou des groupements de population particuliers), ainsi que des lacunes de désagrégation par âge, par sexe ou en fonction d'autres caractéristiques. Les efforts de renforcement des capacités nationales se révèlent essentiels si l'on veut obtenir des résultats de haute qualité en matière de recensement, en particulier dans les pays en développement où i) des recensements n'ont pas été effectués depuis plusieurs décennies, ii) des conflits sociaux et politiques récents ont décimé l'infrastructure nationale, iii) les ressources financières, humaines et d'infrastructure demeurent limitées ou iv) la volonté politique n'est tout simplement pas au rendez-vous.

À cette fin, le Fonds des Nations Unies pour la population (UNFPA) travaille en étroite collaboration avec les pays en développement afin de contribuer à la mise au point de stratégies de défense du déroulement de recensement. Le UNFPA aide aussi les pays à lever des fonds externes pour la tenue de recensements, et il les appuie au niveau de l'élaboration de plans directeurs et de l'implantation d'efforts visant au déroulement de recensements. De plus, grâce à des ateliers de travail, des directives et des manuels

portant sur divers sujets techniques et à une coopération technique directe avec le pays, le Programme mondial pour 2010 de la Division de statistique des Nations Unies (DSNU) s'efforce d'encourager tous les pays du monde à effectuer un recensement au cours du cycle actuel et à se conformer aux Principes et recommandations des Nations Unies pour les recensements de la population et de l'habitat (DSNU, 2008) relatifs au contenu et aux processus de déroulement d'un recensement pour garantir des résultats de qualité.

#### **Étude 4. Les répercussions des révisions des *Perspectives sur la population mondiale* sur le TNS+**

La Division de la population des Nations Unies utilise des données issues de recensements et d'enquêtes-ménages pour obtenir des chiffres de population. Nombreux sont les pays qui effectuent des recensements tous les dix ans, alors que certains autres fournissent des données mises à jour sur la population provenant d'enquêtes-ménages comme l'EDS et l'EGIM. Pour les années intercalaires entre les recensements ou les enquêtes-ménages, la Division de la population des Nations Unies produit des estimations mises à jour en se servant de renseignements externes sur la fécondité, la mortalité et la migration. Tous les deux ans, la Division de la population des Nations Unies rend publiques des révisions aux données les plus récentes sur la population dans *Perspectives sur la population mondiale*. Celles-ci se répercutent sur le dénominateur des indicateurs de l'ISU.

Les révisions prennent en compte toute nouvelle donnée reçue des pays. Même si aucun nouveau recensement ou enquête-ménage n'a été réalisé depuis la révision antérieure, il est possible, par exemple, que des renseignements mis à jour soient disponibles sur la mortalité associée au VIH/sida, et permettent d'améliorer les estimations pour les années intercensitaires. Par conséquent, la Division de la population des Nations Unies révisé les estimations même en ce qui concerne les années pour lesquelles elle ne dispose d'aucuns chiffres de recensement mis à jour. Toutefois, si un nouveau recensement ou enquête-ménage a été réalisé depuis la dernière révision de la Division de la population des Nations Unies, les données sont alors ajoutées et les estimations relatives aux années entre ce recensement (p. ex. l'an 2000) et le recensement antérieur (p. ex. l'an 1990) sont révisées. La plupart du temps, cela permet l'obtention de meilleures estimations par rapport à celles de la révision précédente, puisque la mise à jour comporte une « révision à rebours » comprise entre deux points fixes (les années 1990 et 2000) plutôt qu'une projection vers l'avenir fondée sur un point fixe (l'année 1990). De toute évidence, à mesure que l'on s'éloigne d'une année de recensement, les estimations deviennent moins précises, en particulier lorsqu'il est question de projections vers l'avenir.

L'ISU met à jour nombre de ses indicateurs (incluant le TNS+) en se fondant sur les données révisées sur la population. Il y a lieu de noter que l'ISU ne met à jour que le dénominateur du TNS et du TNS+, de telle sorte que tout changement observé au niveau du TNS et du TNS+ découle des révisions de la population et non des révisions des données d'inscription.

Cette étude compare la différence entre les chiffres du TNS+ fondés sur deux éditions de révisions (2002 et 2004) des *Perspectives sur la population mondiale*. Les différences relatives en pourcentage entre ces données sont fournies au **tableau 9**; les différences de plus de 5 % en valeur absolue sont mises en évidence. Il y a lieu de noter que bien que les estimations du TNS+ provenant du tableau 1 sont fondées sur l'édition de 2004 des révisions, elles s'avèrent différentes de celles du tableau 9 parce qu'il s'agit de taux combinés applicables à de multiples années scolaires. On notera également que le

tableau 9 englobe des pays visés par l'EGIM, le Tchad, le Niger et le Sénégal, étant donné qu'ici l'enquête ne vise que des données de recensement.

Dans certains cas, on peut constater l'existence de différences substantielles entre les estimations issues des deux cycles de recensement comme, par exemple avec le Bangladesh à 9 %. Dans ce cas, l'estimation du TNS+ pour 2000 fondée sur la révision du recensement de 2002 n'a probablement pas pris en compte les données les plus récentes du recensement effectué en 2001 étant que, souvent, il faut compter quelques années pour le traitement des données du recensement et leur envoi à la Division de la population des Nations Unies. Le chiffre de 84,55 % était selon toute probabilité une projection vers l'avenir fondée sur le recensement effectué en 1991, neuf ans avant l'an 2000. Conséquemment, personne ne sera surpris de constater un changement substantiel au niveau du chiffre fondé sur la révision de 2004 (93,09 %), étant donné que ce chiffre avait vraisemblablement pris en compte le recensement de 2001 et qu'une révision à rebours avait été réalisée pour l'année 2000, toute une amélioration si l'on considère que cela s'est fait une année seulement avant la tenue du recensement.

Une situation semblable s'est sans doute produite au Niger et au Sénégal. Il y a lieu de prendre note que, dans le cas du Bangladesh, la révision entre 2002 et 2004 a substantiellement accru la différence entre les chiffres pour le TNS+ et le TNF+ (ce dernier totalisant 79,59 %). Néanmoins, la révision des données de population constitue probablement une amélioration.

***Recommandation 6 :*** *Bien que peu de choses puissent être faites pour remédier à cette situation, les analystes devraient faire preuve de prudence avec les estimations du TNS+ si les dénominateurs sont fondés sur des renseignements de recensement de la Division de la population des Nations Unies qui remontent à plusieurs années. Il est reconnu que les évaluations fondées sur des recensements qui ont été effectués à bonne distance temporelle de l'année scolaire dont il est question sont susceptibles d'être moins précises et peuvent être sources d'erreurs, ce qui expliquerait les écarts entre les chiffres du TNS+ et du TNF+.*

**Tableau 9. Les effets sur le TNS+ des révisions de recensement**

Pays	Année du recensement le plus récent	Année des données d'inscription	TNS+ (révision de recensement 2002)	TNS+ (révision de recensement 2004)	Différence relative en pourcentage (par rapport à 2004)
Bangladesh	2001	2000	84,55	93,09	-9,17
Burkina Faso	1996	2003	35,48	37,28	-4,83
Côte d'Ivoire	1998	1999	56,18	53,34	5,32
Égypte	1996	2000	90,29	93,50	-3,43
Ghana	2000	2003	63,40	62,19	1,95
Guinée	1996	1999	45,05	43,16	4,38
Inde	2001	2000	81,31	80,90	0,51
Indonésie	2000	2003	97,08	101,08	-3,96
Kenya	1999	2003	78,69	77,36	1,72
Mali	1998	2001	.	.	.
Mozambique	1997	2004	72,64	71,04	2,25
Namibie	2001	2000	77,80	74,23	4,81
Niger	2001	2000	26,27	24,49	7,27
Nigéria	1991	2003	62,78	62,09	1,11
Rwanda	2002	2000	.	.	.
Sénégal	2002	2000	59,68	53,46	11,63
Tanzanie	2002	1999	46,37	47,70	-2,79
Tchad	1993	2000	54,94	51,87	5,92
Viet Nam	1999	2002	97,13	96,06	1,11

**Note :** Pour l'Indonésie, le TNS+ pour 2003 fondé sur l'édition 2004 de révisions de recensement est supérieur à 100 % ; cette question est abordée un peu plus loin au chapitre 4.

#### **4. Enquêtes relatives à la combinaison de données d'inscription provenant de sources administratives à des données sur la population issues de recensements**

##### **Étude 5. Les répercussions sur la construction du TNS+ de l'utilisation de deux sources de données distinctes**

Tel qu'évoqué antérieurement, pour construire le TNS+, l'ISU fait appel à deux sources de données. Pour le numérateur, les données sont obtenues auprès de sources administratives nationales, par l'entremise des Ministères de l'Éducation ou de Bureaux nationaux de la statistique, et elles sont fournies à l'ISU dans le cadre de sa cueillette annuelle de données sur l'éducation. Des données agrégées sur les étudiants, les enseignants et les finances sont obtenus de chaque pays du monde. On obtient en particulier des données cumulatives sur les inscriptions par âge et par année d'études, lesquelles sont utilisées à titre de données d'entrée pour le numérateur de l'indicateur du TNS+. La donnée d'entrée pour le dénominateur est fournie à l'ISU par la Division de la population des Nations Unies. Bien qu'en théorie, l'indicateur du TNS+ ne devrait pas excéder 100 %, en pratique, il arrive que ce soit le cas parce que le numérateur et le dénominateur sont obtenus de sources distinctes de données et que des erreurs peuvent en découler.

Une analyse de l'ensemble de la base de données entre 1999 et 2004 (soit une période de cinq ans), pour environ 200 pays auprès desquels l'ISU recueille des données, a démontré l'existence de près de 250 valeurs du TNS+ qui excèdent 100 % (plus ou moins 25 % des 1 000 valeurs) à l'échelle de 65 pays. En réaction aux cas de TNS+ (et de TNS) qui excèdent 100 %, l'ISU a construit des valeurs « plafonnées » inférieures à 100 % par l'entremise d'une formule qui détermine un chiffre à la baisse au prorata en se servant à la fois des taux masculin et féminin. La méthode se présente comme suit : si le TNS ou le TNS+ dépasse 100 % mais sans excéder 105 %, tant les chiffres originaux de TNS et de TNS+ sont simultanément divisés par le plus élevé des six indicateurs suivants : le TNS, le TNS (féminin), le TNS (masculin), le TNS+, le TNS+ (féminin) et le TNS+ (masculin). Si l'une ou l'autre des valeurs originales du TNS ou du TNS+ dépasse 105 %, aucun plafonnement ne survient alors étant donné que les chiffres sont considérés trop peu fiables et les valeurs originales sont supprimées. Cependant, il semble bien qu'avant d'adopter cette solution, il serait sensé de creuser les raisons qui sous-tendent cette situation et de vérifier s'il existe des enjeux aux racines plus profondes avec l'une ou l'autre des sources des données auxquels il y aurait lieu de s'attaquer.

Il y a lieu de noter qu'un TNS+ dépassant 100 % peut ne pas être en soi une source d'inquiétude. En réalité, on doit s'attendre à ce qu'il y ait des erreurs (tant des biais que de l'incertitude / de la variabilité) à la fois pour le numérateur que le dénominateur du TNS+. Il se peut que les deux chiffres demeurent non biaisés tout en contenant une certaine dose d'erreur aléatoire ou de variabilité dans leur mesure. Lorsque la participation scolaire sera virtuellement universelle et que, conséquemment, la vraie valeur du TNS+ équivaudra à presque 100 %, tant le numérateur que le dénominateur formeront des estimations indépendantes de la même valeur. S'il existe une erreur aléatoire, on s'attendrait alors à ce que, dans la moitié des cas, l'estimation d'inscription dépasserait l'estimation de la population (donnant lieu à un TNS+ supérieur à 100 %) et, dans l'autre moitié des cas, c'est son contraire qui s'imposerait comme une vérité. Par conséquent, s'il n'existe que de légères déviations au-dessus de 100 %, en particulier dans les cas où la participation est tenue comme étant virtuellement universelle, les répercussions sur les taux de participation régionaux ou mondiaux se révéleront alors insignifiants. Dans de telles

circonstances, il est possible que cet enjeu ne mérite pas beaucoup d'attention et que le plafonnement constitue une solution acceptable.

Pour illustrer ce point, examinons le cas du TNS+ pour l'Indonésie en 2003. Le **tableau 9** illustre que la valeur fondée sur le cycle 2002 des révisions de recensement équivaut à 97,08 %, alors que la valeur fondée sur cycle de 2004 des révisions totalise 101,08 %. Étant donné que l'ISU n'effectue pas de mise à jour de la valeur du numérateur pour l'indicateur du TNS+, d'une révision de recensement à une autre, il est évident que lorsqu'un chiffre de moins de 100 % augmente à plus de 100 % au cours de cette période, le changement est entièrement attribuable à la mise à jour de la donnée de recensement illustrée par le dénominateur, et il devient clair que le chiffre de l'estimation de la population a diminué entre les révisions consécutives du recensement. Toutefois, cela ne signifie pas nécessairement que les chiffres mis à jour du recensement (la révision de 2004) étaient fautifs, les révisions ultérieures affichant souvent des valeurs améliorées par rapport à celles des révisions antérieures. À n'en point douter, il est probable que le chiffre de la révision de 2004 était fondé sur le recensement réalisé en 2000 en Indonésie, alors que le chiffre de la révision de 2002 était fondé sur le recensement précédent, effectué quelque douze ans auparavant, en 1990. Conséquemment, il est vraisemblable de croire que le chiffre fondé sur la révision de 2004 dépasse les 100% (et seulement légèrement) à cause des erreurs de mesure aléatoires au niveau soit du numérateur soit du dénominateur.

Revenant à l'analyse de la base de données entre 1999 et 2004, nous observons que nombre des 250 valeurs qui dépassent 100 % ne l'excèdent que faiblement, c.-à-d. par moins de 5 %. Il se produit principalement dans les pays où la participation scolaire se veut virtuellement universelle (p. ex. la Belgique, les Bermudes, le Canada, le Danemark, la Finlande, la France, la Région administrative spéciale de Hong Kong en Chine, l'Islande, Israël, l'Italie, le Japon, les Pays-Bas, la Nouvelle-Zélande, la Norvège, l'Espagne, la Suède, la Suisse et le Royaume-Uni) et que, par conséquent, ces cas ne suscitent que peu d'inquiétude.

Cependant, une valeur qui dépasse considérablement les 100 % demeure une source d'inquiétude, puisque la variation du numérateur et du dénominateur ne peuvent, de façon plausible, expliquer le surplus, et il en va de même lorsque nous avons des raisons de mettre en doute le fait que la participation scolaire soit virtuellement universelle. Dans de tels cas, le chiffre >100 % peut signaler la présence d'un grave problème sous-jacent de données qui nécessite un examen plus poussé. De surcroît, s'il existe de tels problèmes au sujet des données, il est possible que cela ne touche pas seulement les pays où le TNS+ dépasse les 100 %, mais également d'autres pays où le TNS+ pourrait être erronément gonflé, mais dans une mesure qui ne ferait pas en sorte que la valeur dépasse les 100 %.

Dans environ 50 des 250 cas, nous avons un TNS+ qui dépasse les 105 % et ce, à l'échelle de 16 pays. Pour plusieurs de ces pays, il n'y a pas lieu de croire que la participation scolaire touche presque les 100 %. De plus, 13 de ces 16 pays possèdent de très petites populations (Anguilla, le Belize, la Grenade, le Guyana, Kiribati, Montserrat, les Antilles néerlandaises, Niue, Saint-Kitts-et-Nevis, les Seychelles, Tokelau, Tonga et Tuvalu) et la Division de la population des Nations Unies ne fournit pas de données pour les pays affichant une population de moins de 100 000 habitants, ce qui est le cas pour 7 de ces 13 pays. Par conséquent, ou bien l'ISU recueille des données sur la population directement auprès d'eux ou il formule des estimations en se servant de divers moyens.

Cependant, les estimations de la population de l'ISU ne sont pas considérées aussi fiables que les données de la Division de la population des Nations Unies pour les pays plus gros. Et pourtant, même lorsque la population des pays n'est pas si faible (moins de 200 000 mais plus de 100 000) et que la Division de la population des Nations Unies fournit des chiffres de la population (4 de ces 13 pays), celles-ci ne sont souvent pas considérées fiables non plus. Par exemple, certains de ces pays ne fournissent pas de données non cumulatives par âge, ce qui permettrait une vérification de la répartition par âge en tant que mesure globale de la qualité du chiffre de la population totale. Dans de tels cas, des dénominateurs potentiellement sous-estimés pourraient contribuer à expliquer les fortes valeurs de TNS+. Des efforts accrus devraient peut-être être consentis de façon à améliorer les chiffres de population pour ces pays plus petits.

Affichant des valeurs de TNS+ entre 105,5 % et 109 %, on retrouve l'Argentine, le Portugal et la Syrie qui comptent des populations importantes (environ 41 millions, 11 millions et 22 millions respectivement). En ce qui a trait à l'Argentine et au Portugal, on pourrait s'attendre à retrouver une inscription frôlant les 100 %. Toutefois, pour ces deux pays, il existe certaines preuves à l'effet que l'on retrouverait une double comptabilisation de certains étudiants au niveau des inscriptions du numérateur. Il se peut que cela soit attribuable au fait que les chiffres nationaux d'inscription soient recueillis à divers moments de l'année pour différentes écoles au sein du pays et, qu'entre-temps, certains étudiants ayant changé d'écoles, ceux-ci soient conséquemment comptés en double. Dans le cas de la Syrie, le TNS se situe très près de 100 % (moins de 102 %), mais le chiffre du TNS+ apparaît dans la fourchette de 107 % à 109 %. Le fait que le TNS dépasse les 100 % peut tout simplement découler d'erreurs aléatoires comme cela a été évoqué plus tôt. Cependant, l'inflation supplémentaire de 5 % à 7 % du TNS+ par rapport au TNS est vraisemblablement attribuable au fait qu'un nombre substantiel d'enfants sous l'âge officiel sont inscrits à l'école secondaire (une comparaison des définitions du TNS et du TNS+ illustrera ce phénomène).

**Recommandation 7 :** *Entre 1999 et 2004, 25 % des valeurs de TNS+ étaient supérieures à 100 %. Cependant, en majorité, ces valeurs ne dépassaient que très faiblement les 100 % et elles se rapportaient principalement à des pays où il y avait lieu de croire que la participation scolaire était presque universelle (100 %). Dans le cas de ces valeurs, il se peut que le chiffre >100 % soit tout simplement le résultat des erreurs aléatoires attribuables au fait que les numérateurs et les dénominateurs proviennent de sources différentes de données. Dans de telles circonstances, il semble raisonnable de faire appel à la méthodologie de l'ISU en vertu de laquelle les chiffres sont plafonnés pour qu'ils ne dépassent pas les 100 %.*

*Toutefois, il existe un nombre considérable de pays à faible population (moins de 200 000 habitants) où les valeurs du TNS+ sont supérieures à 105 % mais où il demeure peu vraisemblable de croire que la participation frise les 100 %. Dans le cas de ces pays, il est probable que les chiffres du dénominateur pour la population posent problème, et l'ISU devrait investir des ressources substantielles et travailler en étroite collaboration avec la Division de la population des Nations Unies et les pays comme tels afin d'identifier la source du problème. En effet, nombreux sont les indicateurs de l'ISU qui dépendent de la cohérence entre les données administratives de l'ISU et celles sur la population de la Division de la population des Nations Unies. Tel que mentionné à la recommandation 5, les problèmes qui portent sur la population présentent, aux yeux de l'ISU, un enjeu si important qu'on se doit de considérer le financement d'un poste dont le détenteur, au sein de l'ISU, s'attaquerait uniquement à ces problèmes.*

## 5. Enquêtes relatives aux données sur la fréquentation provenant des enquêtes-ménages

### Étude 6. De quelle façon la stratification *a posteriori* des éléments de poids des enquête-ménages se répercute-t-elle sur les estimations des totaux ?

Idéalement, la différence relative en pourcentage, en ce qui a trait au nombre d'enfants d'âge scolaire primaire, entre le recensement et les estimations de l'EDS devrait raisonnablement avoisiner le 0 (*voir tableau 4*). Comme les données de l'EDS et de la Division de la population des Nations Unies révèlent essentiellement la même réalité, ils devraient afficher une ressemblance frappante. Cependant, si l'échantillon de l'enquête a été composé de façon telle à ce que la représentativité en fonction de la répartition par âge (ou d'autres attributs) ne soit pas respectée, des différences apparaîtront alors.

Le tableau 4 illustre que quatre pays (l'Indonésie, le Mozambique, la Namibie et le Rwanda) présentent des différences relatives en pourcentage qui varient entre 10 % et 25 % en valeur absolue, ce qui donne à penser qu'il pourrait y avoir un problème au niveau de la représentativité de l'échantillon (le cas du Vietnam a été élucidé dans l'étude 2). Une des façons de corriger une telle situation consiste à faire appel à une technique dite de stratification *a posteriori*, ou étalonnage, dans le cadre de laquelle un ajustement est apporté aux poids d'échantillonnage de telle sorte qu'ils forment la somme des totaux externes de recensement selon l'âge. Bien que la stratification *a posteriori* n'ait à toutes fins pratiques aucun effet sur les estimations du TNF+ puisqu'il s'agit de ratios, elle se répercute sur les estimations des numérateurs du TNF+ (qui sont des évaluations des totaux) pour deux raisons. Premièrement, elle force les données de l'enquête à refléter la répartition globale par âge implicite qui se dégage des chiffres externes du recensement. Deuxièmement, elle force les données de l'enquête à refléter la population totale des enfants d'âge scolaire primaire affichée par les chiffres externes du recensement. C'est la deuxième raison qui risque vraisemblablement de s'accompagner des répercussions les plus considérables sur les estimations du numérateur du TNF+.

Quoique nous ayons été informés que les pays visés par l'EDS procèdent à des vérifications empiriques, afin de s'assurer qu'il existe une cohérence en matière de répartition par âge entre les enquêtes et les sources nationales sur la population, et qu'une stratification *a posteriori* est effectuée lorsque cela est jugé nécessaire (c.-à-d. lorsqu'il y a un manque de cohérence), aucun des pays visés par l'EDS étudiés dans ce document n'a eu recours à la stratification *a posteriori*. Toutefois, nous l'avons fait et cette étude en illustre les résultats. Nous avons ajusté les poids d'échantillonnage originaux en utilisant les chiffres de la population d'âge unique provenant de la Division de la population des Nations Unies de façon à disposer d'un ensemble de poids ajustées « stratifiées *a posteriori* ». Ces poids ont ensuite été utilisés pour recalculer les estimations de l'enquête et pour s'assurer de l'existence d'une exacte correspondance entre les chiffres mentionnés au tableau 4 pour le dénominateur de l'EDS et de la Division de la population des Nations Unies. Les poids stratifiés *a posteriori* ont ensuite été utilisés pour recalculer les estimations fondées sur l'enquête du tableau 3 afin de vérifier la présence d'une amélioration relative du chiffre sur l'inscription. Les résultats de ces nouveaux calculs apparaissent dans les trois dernières colonnes du **tableau 10**.

La différence reste très mince lorsque l'estimation du TNF+ original provenant du tableau 1 (calculé à l'aide des poids « transformés ») est comparée à l'évaluation du TNF+ (calculée à l'aide des poids stratifiés *a posteriori*). On est en droit de s'y attendre puisque les estimations des ratios ne sont pas grandement influencées par une stratification *a posteriori*.

**Tableau 10. Les répercussions de la stratification *a posteriori* sur les estimations**

Pays	TNF+ (original provenant du tableau 1)	TNF+ (obtenu à l'aide des poids stratifiés <i>a posteriori</i> )	Nombre total d'inscriptions d'âge scolaire primaire (chiffre administratif original de l'ISU fondé sur un chiffre combiné)	Nombre total pour la fréquentation d'âge scolaire primaire (estimation de l'EDS à l'aide des poids stratifiés <i>a posteriori</i> )	Différence relative en pourcentage (par rapport aux élèves inscrits)
Bangladesh	79,59	79,34	15 020 499	12 727 738	-15,26
Côte d'Ivoire	52,15	52,63	1 473 852	1 420 720	-3,60
Égypte	85,50	85,73	7 340 000	6 730 628	-8,30
Indonésie	95 25	95 31	25 184 874	23 588 450	-6,34
Mozambique	59 85	61 03	2 317 869	2 242 635	-3,25
Namibie	78,60	78,30	283 802	297 963	4,99
Nigéria	61,99	63,09	13 211 302	13 299 210	0,67
Rwanda	71,94	71,97	1 046 634	1 057 650	1,05
Tanzanie	53,83	53,70	3 105 435	3 491 601	12,44
Viet Nam	96,27	96,05	8 498 039	8 494 682	-0,04

Toutefois, les estimations des totaux (tel qu'illustré dans les trois dernières colonnes du tableau 10) en ressortent grandement influencées. La différence relative en pourcentage pour tous les pays est considérablement diminuée par comparaison à celle du tableau 3, alors que seuls le Bangladesh (-15,26 %) et la Tanzanie (12,44 %) affichent des chiffres supérieurs à 10 % en valeur absolue. Dans ce dernier cas, la stratification *a posteriori* semble avoir empiré la situation en comparaison avec le tableau 3. Tel que mentionné précédemment, l'écart pour le Bangladesh est vraisemblablement associé à la façon dont la question portant sur la fréquentation a été posée (*voir Étude 9*), alors on ne s'attend pas à ce que la stratification *a posteriori* améliore la situation. Cependant, pour ce qui est de la Côte d'Ivoire, de l'Égypte, de l'Indonésie, du Mozambique, de la Namibie, du Nigéria et du Rwanda, les différences relatives fortes en pourcentage du tableau 3 ont été radicalement diminuées.

Idéalement, les estimations de l'enquête devraient être stratifiées *a posteriori* en fonction des chiffres « étalonnés » du recensement fournies par les pays (c.-à-d. les chiffres nationaux) et non ceux de la Division de la population des Nations Unies. Dans ce rapport, les chiffres de la Division de la population des Nations Unies sont principalement utilisés pour illustrer l'amélioration éventuelle pouvant être apportée aux estimations des totaux et attribuables au manque d'accès à des chiffres nationaux étalonnés du recensement. Il faut éviter de forcer les données de l'enquête à se conformer aux éléments de référence (*benchmarks*) de la Division de la population des Nations Unies parce que cela viendra déformer les chiffres de la Division de la population des Nations Unies, en tant que source indépendante, par comparaison à la mesure administrative du TNS+ qui se sert des chiffres de la Division de la population des Nations Unies. Si le chiffre du TNF+ est déformé, il devient beaucoup plus difficile de savoir exactement à quel moment le chiffre de TNS+ est fiable et à quel moment il ne l'est pas.

**Recommandation 8 :** En plus de donner accès aux poids originaux (recommandation 3), tant l'EDS que l'EGIM devraient envisager de stratifier *a posteriori* leurs poids originaux pour tous les pays par année d'âge (et peut-être d'autres attributs par la même occasion). Elles devraient utiliser des sources nationales comme points de référence avant de les diffuser sous forme de fichiers de micro-données aux fins d'une utilisation publique. Cela

améliorera les estimations des totaux, car les poids rendront les estimations conformes aux contrôles externes de recensement des sources nationales. Par conséquent, cela corrigera tout écart, eu égard à la représentativité de l'échantillon par rapport aux attributs qui sont des points de référence. Il n'est pas recommandé de stratifier a posteriori les chiffres de la Division de la population des Nations Unies. Pour plus de détails sur la stratification a posteriori, consultez Sarndal et al., 1992.

### **Étude 7. Erreurs non liées à l'échantillonnage : un aperçu de la non-réponse des enquête-ménages au niveau de l'EDS**

Il existe de nombreuses sources éventuelles d'erreurs non liées à l'échantillonnage dans le cadre des enquête-ménages, incluant les erreurs de non-réponse, les erreurs de couverture de la base de sondage, les erreurs associées aux instruments (questionnaires), les erreurs de l'intervieweur, les erreurs du répondant, les erreurs de traitement, les erreurs d'estimation et les erreurs d'analyse. Bien que le dépistage des erreurs d'échantillonnage dans le cadre des enquêtes fasse souvent l'objet d'une attention exagérée, les erreurs non liées à l'échantillonnage passent fréquemment inaperçues et ce, même si celles-ci peuvent être plus considérables que les premières. La plupart du temps, les commanditaires des enquêtes n'investissent pas suffisamment de ressources au niveau du dépistage et du contrôle des sources d'erreurs non liées à l'échantillonnage parce que cela se révèle à la fois onéreux et difficile à faire. Par exemple, une des façons de déceler un biais attribuable à une non-réponse d'enquête ménage consiste à mettre en œuvre une étude de suivi de non-réponse, dans le cadre de laquelle on entre à nouveau en contact avec un sous-échantillon de non-répondants à une enquête et on effectue une tentative de les interviewer une fois que l'enquête principale a pris fin afin de vérifier si leurs réponses à l'enquête diffèrent systématiquement de celles des répondants à l'enquête principale. Il s'agit là de toute évidence d'une tâche dispendieuse qui est rarement menée à bien.

Toutefois, en ce qui concerne l'une des sources des erreurs non liées à l'échantillonnage, la non-réponse, il est facile de la dépister en examinant l'ensemble des taux de réponse, au niveau du ménage et des items individuels. L'étude 7 portera sur cette question. Si l'on néglige de s'attaquer aux taux élevés de non-réponse, il peut en découler des problèmes en ce qui a trait à la qualité des données. Si le taux de non-réponse reste modeste, on peut possiblement résoudre la question en réévaluant les poids des répondants par un haussement de leur poids d'échantillonnage, ce qui prendra en compte les non-répondants qui ne se retrouvent pas dans l'échantillonnage. Cette méthode présume de façon implicite que, si les non-répondants avaient répondu à l'enquête, ils auraient fourni des données similaires à celles des répondants (ce que l'on désigne par l'expression « manquant complètement de façon aléatoire » ou la présomption *MCAR* [*missing completely at random*]). Bien que l'on sache que cela n'est pas souvent le cas et qu'en recourant à la présomption *MCAR*, on ouvre la porte à une certaine dose de distorsion, il n'y a pas grand-chose d'autre à faire pour remédier à la situation sans en savoir davantage au sujet de l'étendue réelle du biais. Souvent, les méthodes améliorées comprennent l'ajustement des poids au sein des catégories de réponse (p. ex. les régions géographiques), là où l'on peut présumer que les non-répondants et les répondants demeurent plus susceptibles de donner des réponses semblables (ce que l'on désigne par l'expression « manquant de façon aléatoire » ou la présomption *MAR* [*missing at random*]), qui reste plus modérée que la présomption *MCAR*).

Pour tous les 16 pays, les taux de réponse soumis pour l'EDS au niveau des ménages restent très élevés (>95 %) (voir **tableau 11**) et, par conséquent, ils ne sont pas considérés problématiques.

**Tableau 11. Taux de réponse des ménages pour l'EDS**

Pays	Taux de réponse des ménages
Bangladesh	99,3
Burkina Faso	99,4
Côte d'Ivoire	97,9
Égypte	99,1
Ghana	98,7
Guinée	97,3
Inde	-
Indonésie	98,9
Kenya	96,3
Mali	98,0
Mozambique	-
Namibie	96,9
Nigéria	98,6
Rwanda	99,5
Tanzanie	98,3
Viet Nam	99,9

Il est tout aussi important de vérifier le taux de réponse par item, à savoir le taux de réponse par rapport à la variable précise utilisée pour mesurer la fréquentation.

Le taux de non-réponse par item peut être défini comme étant=

$$100 \frac{\text{Nombre denfants d'âge scolaire primaire échantillonnés (non pondéré) qui ont répondu à la question sur la fréquentation}}{\text{Nombre denfants d'âge scolaire primaire échantillonnés (non pondéré) à qui a été posée la question sur la fréquentation}}$$

Les résultats de cette analyse apparaissent au **tableau 12**, et ils indiquent que les taux de réponse par item étaient également très élevés (>97 %) pour l'ensemble des 16 pays et que, conséquemment, ils ne sont pas considérés problématiques non plus.

**Tableau 12. Taux de réponse pour l'item sur la fréquentation (fréquentation de l'école au cours de l'année scolaire actuelle/toujours à l'école) pour l'EDS**

Pays	Taux de réponse par item
Bangladesh	99,43
Burkina Faso	99,89
Côte d'Ivoire	97,72
Égypte	100,00
Ghana	99,85
Guinée	99,77
Inde	99,60
Indonésie	99,48
Kenya	99,92
Mali	99,84
Mozambique	99,89
Namibie	99,44
Nigéria	99,75
Rwanda	99,83
Tanzanie	99,90
Viet Nam	99,97

## Étude 8. Erreurs d'échantillonnage : Un aperçu des erreurs-types et des coefficients de variation pour l'EDS

Les erreurs d'échantillonnage élevées demeurent une source éventuelle de problèmes quant à la qualité des données. Par erreur d'échantillonnage, on entend la différence entre l'estimation découlant d'une enquête auprès d'un échantillon et la vraie valeur qui serait obtenue si un recensement de l'ensemble de la population était entrepris en vertu des mêmes conditions. Une des façons de mesurer l'étendue d'une erreur d'échantillonnage consiste à calculer l'erreur-type de l'estimation. Pour y arriver, on peut recourir à une variété de techniques, incluant la linéarisation de Taylor et l'application de la méthode *jackknife* (voir *Wolter, 1995*). Le pourcentage de coefficient de variation (CV) demeure communément utilisé, ce qui suppose la division de l'erreur-type par l'estimation, le tout multiplié par 100. Souvent, c'est le CV que l'on présente plutôt que l'erreur-type. Il en est ainsi parce que le CV constitue une quantité sans échelle qui permet les comparaisons, à la fois entre les estimations de même type (tels que les TNF+ d'un pays à un autre) et les estimations qui comportent diverses unités de mesure (ce qui n'est pas le cas dans cette étude).

Dans le **tableau 13**, l'erreur-type du TNF+ a été calculée en se servant à la fois de la méthode de linéarisation de Taylor (grâce au progiciel SAS) et à la méthode *jackknife* (grâce à WesVar, un gratuiciel téléchargeable à partir de Westat). Ces deux méthodes devraient donner des résultats similaires et c'est le cas en tout temps. Il y a lieu de noter que le TNF+ et les erreurs-types correspondantes dans le tableau 13 ont été calculés en faisant appel à des poids « transformés » (voir *Étude 2 pour une description des poids « transformés »*). Un calcul similaire a aussi été effectué en se servant d'éléments de poids stratifiés *a posteriori*, mais cela a donné une très faible différence au niveau soit des estimations ponctuelles ou des erreurs-types par rapport à ce qui est rapporté au tableau 13.

Les CV apparaissent aussi dans le tableau 13. Ils sont tous inférieurs à 7 % (et souvent bien plus bas que 7 %), ce qui indique la présence d'estimations très précises du TNF+. En général, si les CV sont inférieurs à 16 %, les estimations ponctuelles correspondantes sont considérées publiables; par contre, les estimations ponctuelles affichant des CV variant de 16 % à 33 % ne devraient être publiées qu'avec réserve et celles qui présentent des CV supérieurs à 33 % devraient être supprimées. Selon cette norme, toutes les estimations du TNF+ dans le tableau 13 sont publiables et, de toute évidence, les erreurs d'échantillonnage ne constituent pas un problème important dans ces pays.

**Tableau 13. Erreur-types et coefficients de variation (CV) pour l'EDS**

Pays	TNF+	Erreur-type du TNF+ (Taylor)	Erreur-type du TNF+ (Jackknife)	Pourcentage du coefficient de variation (utilisant le Jackknife)
Bangladesh	79,59	0,82	0,822	1,032
Burkina Faso	32,33	1,56	1,561	4,828
Côte d'Ivoire	52,15	2,70	2,698	5,173
Égypte	85,50	0,57	0,567	0,663
Ghana	61,22	1,24	1,240	2,025
Inde	82,33	0,35	0,351	0,426
Indonésie	95,25	0,32	0,345	0,362
Kenya	74,99	1,02	1,022	1,362
Mali	39,03	1,42	1,421	3,640
Mozambique	59,85	1,03	1,030	1,720
Namibie	78,60	0,93	0,930	1,183
Nigéria	61,99	1,69	1,693	2,731
Rwanda	71,94	0,72	0,724	1,006
Tanzanie	53,83	1,70	1,705	3,203
Viet Nam	96,27	0,59	0,627	0,651

On notera avec intérêt que les erreurs-types et les CV du TNF+ sont peu élevés à cause de la forte corrélation entre le numérateur et le dénominateur, étant donné qu'il s'agit dans les deux cas d'estimations fondées sur l'enquête. Sur le plan de la théorie, en l'absence de tout biais, l'estimation de l'enquête peut se révéler préférable à une estimation découlant de sources administratives tout simplement parce que, dans ce dernier cas, les valeurs utilisées pour le numérateur et le dénominateur ne sont pas corrélées. Par conséquent, étant donné la grande précision des estimations mises en évidence dans le tableau 13, la question de savoir si oui ou non l'enquête devrait être considérée comme une source de données privilégiée est reliée à l'étendue et à la profondeur de tout biais éventuel (p. ex. la non-réponse, les erreurs de mesure, la non-couverture issue du plan d'échantillonnage global ou de l'omission d'une population n'appartenant pas à des ménages). La première de celles-ci (la non-réponse) a fait l'objet d'une enquête dans l'étude 7; un examen complet des deux dernières (les erreurs de mesure et la couverture) outrepasserait cependant le mandat de ce rapport.

Un commentaire final s'impose au sujet du traitement des valeurs manquantes. Partout dans ce rapport, les valeurs du TNF+ ont été calculées en omettant les valeurs manquantes du fichier des données de l'enquête; aucune imputation des valeurs manquantes n'a été réalisée. Les erreurs-types ont été calculées de la même façon. Bien que cela soit loin d'être idéal étant donné les possibilités de biais, le nombre de valeurs manquantes pour la plupart des pays de l'EDS était si faible que leur omission ne s'accompagnerait vraisemblablement que de répercussions négligeables sur la valeur calculée.

## Étude 9. L'impact des modalités de saisie de la fréquentation de l'EDS dans divers pays

Dans le contexte de l'EDS, on réussit généralement à saisir la fréquentation à l'école primaire grâce à deux questions : l'une a trait à la fréquentation courante et l'autre portant sur la fréquentation (durant une journée ou plus) en tout temps au cours de l'année scolaire actuelle (s'il n'y a pas fréquentation sur le moment). La dernière question vise à saisir les enfants qui ne fréquentent pas l'école pour cause de motifs temporaires tels que la maladie. Afin d'évaluer le nombre d'enfants non scolarisés, l'ISU et l'UNICEF travaillent à la mise au point d'une méthodologie qui utiliserait conjointement les données administratives et celles de l'enquête. En ce qui concerne la composante de l'enquête, la définition de la « scolarisation » est dépitée par l'entremise de la notion de fréquentation. On considère qu'un enfant fréquente l'école s'il a répondu oui à l'une ou l'autre des deux questions posées par l'EDS. Toutefois, bien que les deux questions soient posées dans certains pays (voir **tableau 14**), dans d'autres, on ne pose que la question relative à la fréquentation courante (voir **tableau 15**). La plupart du temps, la question sur la fréquentation courante saisie la portion la plus importante de la fréquentation scolaire bien que cela ne soit pas toujours le cas.

**Tableau 14. Comparaison entre le TNS+ et le TNF+ pour les pays où tant la question sur la fréquentation courante que celle sur la fréquentation passée ont été posées dans le cadre de l'EDS**

Pays	Fréquentation courante	Fréquentation passée	Fréquentation courante + passée : TNF+ (B)	TNS+ combiné (A)	Différence (A-B)
Burkina Faso	31,85	0,47	32,32	38,6	6,28
Égypte	85,42	0,07	85,49	93,5	8,01
Ghana	61,06	0,14	61,20	59,8	-1,40
Guinée	27,46	12,49	39,95	43,16	3,21
Kenya	74,87	0,11	74,98	77,36	2,38
Mali	38,96	0,07	39,03	42,98	3,95
Mozambique	58,78	1,07	59,85	62,91	3,06
Namibie	78,31	0,29	78,60	74,23	-4,37
Nigéria	61,59	0,40	61,99	62,09	0,10
Rwanda	43,74	28,19	71,93	71,10	-0,83
Tanzanie	53,70	0,13	53,83	47,70	-6,13

**Tableau 15. Comparaison entre le TNS+ et le TNF+ pour les pays où seule la question sur la fréquentation courante a été posée dans le cadre de l'EDS**

Pays	Fréquentation courante : TNF+ (B)	TNS+ combiné (A)	Différence (A-B)
Bangladesh	79,34	93,05	13,71
Côte d'Ivoire	52,63	53,67	1,04
Inde	82,51	82,34	-0,17
Indonésie	95,31	100,86	5,55
Viet Nam	96,05	96,06	0,01

Par exemple, le tableau 14 montre que 28 % de la fréquentation totale de l'école primaire au Rwanda (72 %) et que 12,5 % de celle de la Guinée (40 %) ont été saisies par la deuxième question. Par comparaison aux autres pays, ce sont là des chiffres exceptionnels. Il s'avère difficile de savoir si ces résultats représentent la réalité en ce qui a trait à la fréquentation (ces résultats pourraient découler d'un bouleversement important des services d'enseignement durant l'année à cause de la sécheresse ou de la guerre civile) ou s'ils représentent dans les faits une grave divergence quelconque au niveau de la façon dont l'enquête a été effectuée au Rwanda et en Guinée par rapport aux autres pays. Dans ce dernier cas, la divergence reste attribuable à un artefact de l'enquête plutôt qu'à un vrai phénomène de fréquentation.

Le moment choisi pour la tenue de l'enquête dans le contexte de l'année scolaire pourrait constituer l'une des explications possibles. Au Rwanda, l'enquête s'est déroulée de juin à décembre (*voir tableau 1*), alors que l'année scolaire débute en septembre pour se terminer en juin (*voir tableau 2*). Alors que l'enquête se déroulait sur le terrain durant les mois de juillet et août (deux des sept mois de l'enquête), les enfants profitaient des vacances scolaires. Étant donné cet artefact chronologique, il demeure possible que la question portant sur la fréquentation courante posée durant ces deux mois puisse avoir suscité plus de réponses négatives et que la question portant sur la fréquentation passée puisse avoir généré plus de réponses positives que ce à quoi on se serait normalement attendu. Dans les faits, la différence entre le TNF+ total (calculé à l'aide du poids « transformé », colonne B) et le TNS+ (colonne A) demeure si minime pour le Rwanda (<1 % en valeur absolue) qu'il y a fort à parier que le TNF+ total est correct et que, conséquemment, l'explication ci-dessus s'impose comme étant plausible dans le cas du Rwanda.

Un artefact chronologique semblable s'est vraisemblablement produit en Guinée, alors que l'enquête s'y est déroulée d'avril à août (*voir tableau 1*), mais où l'année scolaire débute en octobre pour se terminer en juin (*voir tableau 2*). Lors du déroulement de l'enquête sur le terrain au cours des mois de juillet et août (deux des cinq mois de l'enquête), les enfants profitaient des vacances scolaires. Il n'en demeure pas moins que la curiosité reste sollicitée par le fait que, dans trois autres pays du tableau 14, l'enquête s'est déroulée durant des mois de vacances (au Burkina Faso durant trois des sept mois de l'enquête, au Ghana durant un des cinq mois de l'enquête et au Nigéria durant un des six mois de l'enquête), mais le moment choisi pour effectuer l'enquête n'a pas semblé susciter les mêmes répercussions, pas même dans une mesure perceptible, contrairement à ce qui a été le cas au Rwanda et en Guinée. Cet écart peut illustrer des divergences au niveau des mots utilisés lors de l'entrevue dans les différents pays.

Lorsque l'on prend en considération les pays où seule la question sur la fréquentation courante a été posée (*voir tableau 15*), le Bangladesh se distingue par une différence substantielle entre le TNF+ total (colonne B) et le TNS+ (colonne A). Cette différence (13,71 %) peut s'expliquer par le fait que les données sur la fréquentation passée n'ont pas été saisies. Aucun artefact chronologique d'enquête n'a été observé dans ce pays contrairement au Rwanda et à la Guinée; au Bangladesh, l'année scolaire englobe la totalité des douze mois de l'année. Étant donné que les renseignements manquants au sujet de la fréquentation passée peuvent avoir comptés pour une portion significative de la fréquentation dans ce pays, il est possible que le chiffre d'inscription pour le Bangladesh puisse être considéré comme plus crédible que celui sur la fréquentation.

Il y a aussi lieu de noter que bien que l'on puisse observer une divergence modeste pour l'Indonésie (5,55 %) dans le tableau 15, les tableaux 3 et 4 révèlent l'existence, dans ce pays, d'un problème tant au niveau des composantes du numérateur que du dénominateur. Par conséquent, le problème observé en Indonésie ne peut être pleinement expliqué par la façon dont la question sur la fréquentation a été posée, car cela n'influence que le numérateur. On notera également avec intérêt que dans l'étude 4, le TNS+ fondé sur la révision du recensement de 2002 pour l'Indonésie révélait une valeur (97,08 %) considérablement plus proche du TNF+ du tableau 15 que de la valeur du TNS+ dans le tableau 15.

**Recommandation 9 :** *Tous les pays participant à une EDS devraient envisager de poser les deux questions ayant trait à la fréquentation scolaire (courante et passée) afin de s'assurer que l'ensemble du phénomène de la participation scolaire soit pleinement saisi et de faire en sorte que l'indicateur du TNF+ se compare autant que faire se peut au TNS+. Les intervieweurs devraient être formés à expliquer aux répondants le sens exact de « fréquentation courante » et de « fréquentation passée », en particulier lorsque l'enquête se déroule durant des mois de vacances. Sinon, le questionnaire de l'EDS pourrait être modifié pour prendre en compte les mois de vacances.*

## 6. Enquêtes relatives à la déclaration de l'âge à l'échelle des trois sources de données (enquête, administrative et recensement)

### Étude 10. L'impact de la façon dont l'âge est saisi en matière de fréquentation par opposition à l'inscription

En réexaminant le tableau 1, nous avons noté avec intérêt que la valeur combinée du TNS+ est presque toujours supérieure à la valeur du TNF+. Une explication plausible consiste à dire que certains enfants s'inscrivent à l'école sans jamais la fréquenter. Toutefois, comme la définition du TNS+ et du TNF+ englobe les enfants d'âge scolaire primaire dans le numérateur, il demeure possible que la source administrative (à partir de laquelle le TNS+ est calculé) impute de façon systématique à un nombre supérieur d'enfants la caractéristique d'être d'âge scolaire primaire (soit correctement ou incorrectement) que l'enquête-ménage (à partir de laquelle le TNF+ est calculé). Par conséquent, ce n'est pas tout simplement ici le fait de savoir si l'inscription ou la fréquentation sont saisies correctement qui constitue l'enjeu, mais bien plutôt la façon dont l'âge est capté par les deux instruments.

Afin d'enquêter sur l'ensemble de cette question, une étude a été réalisée sur les modalités de saisie de l'âge par chacune de ces deux sources. Trois proportions ont été calculées.

Le pourcentage d'enfant d'âge scolaire exact =

$$100 \frac{\text{Nombre d'enfants d'âge scolaire primaire inscrits à (fréquentant) l'école primaire}}{\text{Nombre d'enfants de tout âge inscrits à (fréquentant) l'école primaire}}$$

Le pourcentage d'enfants au-delà de l'âge scolaire (trop vieux) =

$$100 \frac{\text{Nombre d'enfants au-delà de l'âge scolaire primaire inscrits à (fréquentant) l'école primaire}}{\text{Nombre d'enfants de tout âge inscrits à (fréquentant) l'école primaire}}$$

Le pourcentage d'enfants en deçà de l'âge scolaire (trop jeunes) =

$$100 \frac{\text{Nombre d'enfants en deçà de l'âge scolaire primaire inscrits à (fréquentant) l'école primaire}}{\text{Nombre d'enfants de tout âge inscrits à (fréquentant) l'école primaire}}$$

Le **tableau 16** présente à la fois des données d'inscription et la fréquentation de l'école primaire. Il y a lieu de noter que l'addition des trois proportions relatives aux inscriptions totalise 100 %, tout comme c'est le cas pour la fréquentation. La définition de l'âge scolaire primaire apparaît au **tableau 17** pour les 16 pays en question visés par l'EDS.

Dans le tableau 16, aucun modèle apparent n'émerge quant aux élèves qui sont trop jeunes pour être inscrits ou pour fréquenter l'enseignement primaire. Toutefois, le « pourcentage d'âge scolaire exact » reste systématiquement plus élevé pour l'inscription que pour la fréquentation (sauf en ce qui concerne l'Indonésie et le Nigéria). Dans le même ordre d'idées, la catégorie des élèves qui sont trop vieux pour le niveau primaire demeure systématiquement plus basse en ce qui a trait à l'inscription par rapport à la

fréquentation (à l'exception des deux mêmes pays). De deux choses l'une, cela signifie que les données de source administrative indiquent incorrectement la présence d'un trop grand nombre d'enfants d'âge exact à l'école primaire ou que la source d'enquête-ménage les compte incorrectement comme étant d'un nombre pas suffisamment élevé. De surcroît, la taille de l'écart entre les données sur la fréquentation et sur l'inscription demeure relativement importante (dans plus de la moitié des pays, l'écart atteint ou est supérieur à 9 %), ce qui donne à penser qu'il existe une certaine forme de sérieuse incohérence entre la façon dont les données sur l'âge sont recueillies et utilisées par l'enquête et les sources administratives.

**Tableau 16. Comparaison des modalités de saisie de l'âge pour les données relatives à l'inscription et à la fréquentation**

Pays	Pourcentage trop jeune		Pourcentage d'âge exact		Pourcentage trop vieux	
	Fréquentation	Inscription	Fréquentation	Inscription	Fréquentation	Inscription
Bangladesh	5,92	7,14	65,99	81,82	28,09	11,04
Burkina Faso	5,42	12,29	74,21	77,59	20,37	10,11
Côte d'Ivoire	.	7,46	69,05	75,91	30,95	16,63
Égypte	0,17	1,17	83,40	92,36	16,43	6,47
Ghana	0,92	2,70	62,75	73,78	36,33	23,52
Inde	.	9,09	75,14	84,04	24,86	6,87
Indonésie	11,06	10,20	83,60	82,76	5,34	7,04
Kenya	1,29	1,07	50,26	69,39	48,46	29,54
Mali	6,31	.	72,94	.	20,74	.
Mozambique	0,79	0,00	62,34	74,83	36,87	25,17
Namibie	0,48	0,04	70,92	72,81	28,60	27,15
Nigéria	3,47	7,09	65,95	57,15	30,58	35,76
Rwanda	2,53	4,98	68,43	82,30	29,04	12,71
Tanzanie	4,54	0,11	69,81	74,07	25,65	25,82
Viet Nam	2,96	0,10	72,28	90,88	24,76	9,02

**Tableau 17. Définition de l'âge scolaire primaire par pays**

Pays	Âge au début	Âge à la fin	Durée (années)
Bangladesh	6	10	5
Burkina Faso	7	12	6
Côte d'Ivoire	6	11	6
Égypte	6	10	5
Ghana	6	11	6
Guinée	7	12	6
Inde	6	10	5
Indonésie	7	12	6
Kenya	6	11	6
Mali	7	12	6
Mozambique	6	12	7
Namibie	6	12	7
Nigéria	6	11	6
Rwanda	7	12	6
Tanzanie	7	13	7
Viet Nam	6	10	5

L'une des explications possibles des résultats qui sont présentés au tableau 16 serait que les données de source administrative sur l'inscription à l'école primaire englobent des enfants surâgés. Pour vérifier cette hypothèse, on peut opter pour la production d'un facsimilé fondé sur l'enquête de ce à quoi ressemblerait le taux apparent de participation si les enfants surâgés fréquentant l'école primaire étaient à tort inclus dans l'évaluation d'inscription. Par exemple, si tous les enfants (peu importe leur âge) de l'école primaire étaient inclus dans le taux de participation à l'enseignement primaire et que cela faisait en sorte que le taux de participation du facsimilé fondé sur l'enquête correspondait aux taux actuel du TNS, cela viendrait alors appuyer les prétentions à l'effet que le dénombrement d'inscription est faussé par l'inclusion fautive d'enfants surâgés au chiffre d'enfants à l'école primaire. Cela peut être approfondi en comparant le TNS avec le taux brut de fréquentation (TBF) pour voir s'ils sont identiques. Cela suppose la comparaison du TNS, tel qu'évalué par les données administratives, qui est exprimé comme étant :

$$\text{TNS} = 100 \frac{\text{Nombre d'enfants d'âge scolaire primaire inscrits à l'école primaire}}{\text{Nombre d'enfants d'âge scolaire primaire}}$$

avec le TBF tel qu'évalué par les données d'enquête et exprimé comme étant :

$$\text{TBF} = 100 \frac{\text{Nombre d'enfants de tout âge fréquentant l'école primaire}}{\text{Nombre d'enfants d'âge scolaire primaire}}$$

Par définition, le TBF peut dépasser les 100 % comme en témoignent plusieurs pays dans le **tableau 18**. Lorsque l'on compare le TNS et le TBF dans le tableau 18, il est apparent qu'aucun pays ne présente un TNS se rapprochant du TBF.

**Tableau 18. Comparaison du TBF et du TNS afin d'enquêter sur une possible attribution fautive de l'âge par les sources de données administratives**

Country	TBF	TNS (révisé, 2002)
Bangladesh	116,99	89,45
Burkina Faso	42,54	38,22
Côte d'Ivoire	74,29	52,82
Égypte	101,88	93,50
Ghana	95,86	58,96
Inde	108,70	82,34
Indonésie	105,16	96,28
Kenya	143,36	76,69
Mali	52,30	42,98
Mozambique	90,10	62,89
Namibie	107,43	74,08
Nigéria	89,04	60,26
Rwanda	104,08	71,10
Tanzanie	75,04	47,67
Viet Nam	133,35	95,92

Bien que le tableau 18 ne jette aucune lumière supplémentaire sur cette question, dans l'ensemble, cette étude en reste une d'importance susceptible d'offrir une explication plausible pour les écarts entre le TNS+ et le TNF+. Par conséquent, il vaut la peine d'approfondir autant que possible cette enquête. Pour un pays donné, il se révélerait fort

valable d'effectuer un projet-pilote, sur une petite échelle, afin de recueillir à la fois des données d'enquête et celles sur l'inscription et de les faire concorder à une échelle locale. La sélection de quelques villages situés dans la zone de chalandise des écoles et le raccordement d'indices des données d'enquête provenant de ces villages aux données d'inscription (et aux enfants qui n'étaient pas inscrits) offriraient d'extraordinaires aperçus perceptuels.

**Recommandation 10 :** *Les données sur l'âge demeure un sujet si important que des études supplémentaires doivent être réalisées pour mieux comprendre l'enjeu. Un projet pilote sur une petite échelle pour recueillir à la fois des données d'enquête et des données administratives et les faire concorder à une échelle locale s'avérerait fort valable.*

Une autre explication possible pour les résultats présentés au tableau 16 est associée aux données sur la fréquentation. Dans le cadre du module qui pose une question sur la fréquentation, l'EDS ne recueille pas de date de naissance en soi (pour calculer directement l'âge). Le répondant du ménage déclare plutôt l'âge de l'enfant lors de l'enquête en années seulement (plutôt qu'en années et en mois). Il en est ainsi parce que l'EDS n'a pas été conçue comme une enquête sur l'éducation où l'âge constitue une variable d'importance cruciale essentielle à l'analyse. Étant donné le moment choisi pour l'interview de l'enquête, il est possible que l'enfant ait célébré son anniversaire depuis le début de l'année scolaire mais avant que n'ait lieu l'interview de l'enquête sur la fréquentation scolaire. Dans de tels cas, l'EDS pourrait qualifier certains enfants comme étant âgés d'un an de plus que leur âge réel au début de l'année scolaire. Conséquemment, si ces enfants sont sur la ligne de démarcation entre l'âge scolaire primaire et l'âge scolaire secondaire, il se peut qu'ils soient classés comme surâgés pour l'école primaire plutôt que d'être classés comme ayant le niveau d'âge exact. Cela pourrait expliquer en partie les raisons pour lesquelles les estimations du « pourcentage d'enfants au-delà de l'âge scolaire » de l'EDS sont supérieures à celles qui sont fondées sur les données administratives de l'ISU, bien que cela ne puisse fournir une explication en bonne et due forme puisque les écarts demeurent trop prononcés, étant donné que seule la dernière année de l'école primaire est touchée.

**Recommandation 11 :** *Lors de futurs cycles, les enquêtes de l'EDS devraient envisager l'ajout d'une question au module d'éducation qui permettrait la saisie de l'âge de l'enfant en années et mois (ou autrement la date de naissance), afin de permettre la production d'estimateurs de TNF+ plus précis.*

Il se révèle utile d'expliquer ce que l'on entend par l'expression « âge scolaire primaire ». Les pays, sur une base individuelle, définissent cette expression de façons fort diverses. Par exemple, supposons que l'année scolaire dans un pays se déroule de septembre à août et qu'elle recoupe des âges variant de 5 à 10 ans inclusivement. Dans certains pays, on permet aux enfants de s'inscrire à l'école primaire, et conséquemment de devenir d'âge scolaire primaire, s'ils ont dépassé le seuil minimal (cinq ans dans l'exemple ci-dessus) au premier mois de l'année scolaire (septembre dans l'exemple ci-dessus). En général, cette définition ne pose aucun problème étant donné que l'âge scolaire primaire demeure conforme à l'année scolaire.

Cependant, dans d'autres pays, on permet aux enfants de s'inscrire à l'école primaire s'ils franchissent le seuil minimal à un moment quelconque de l'année scolaire. Dans un tel cas, quelques enfants seront âgés de quatre ans au premier jour d'école, et ils célébreront leur cinquième anniversaire à un moment quelconque durant leur première année d'études primaires. Comme l'ISU demande aux pays de soumettre l'âge de l'enfant au

début de l'année scolaire, les enfants qui ont quatre ans à ce moment apparaîtront, selon les données de l'ISU, avoir été inscrits à un âge inférieur, bien que cela ne soit pas le cas. Ayant amorcé une enquête sur ce problème, l'ISU a entrepris de recueillir des renseignements sur la façon dont les différents pays définissent l'âge scolaire primaire mais, à ce jour, aucun ajustement n'a été apporté pour prendre en compte ces écarts.

On notera avec intérêt que le problème de l'âge s'impose également dans le cas des données d'enquête, étant donné qu'elles consignent l'âge de l'enfant au moment où l'enquête est réalisée (et sans qu'il y ait nécessairement un rapport avec l'année scolaire). Ainsi, dans l'exemple ci-dessus, si un enfant est âgé de quatre ans lorsqu'une enquête est effectuée en septembre et qu'il fréquente l'école primaire à ce moment-là, il apparaîtra comme un enfant trop jeune.

Un autre problème ayant trait à l'âge est associé aux estimations de la Division de la population des Nations Unies qui sont utilisées pour les dénominateurs du TNS+. Dans le cas de ces chiffres de population, une cohorte d'âge unique représente des estimations moyennes de tous les enfants ayant atteint un anniversaire particulier au 30 juin. Par conséquent, les estimations démographiques ne sont pas cohérentes avec la définition de l'âge d'inscription si, par exemple, l'année scolaire débute en janvier. L'estimation de la population d'âge scolaire primaire différera de plusieurs points de pourcentage advenant que l'on utilise simplement des estimations démographiques pour les personnes âgées de cinq à dix ans (comme dans l'exemple ci-dessus) et qu'on les compare à des estimations, en ajoutant six mois à l'âge de quatre ans, les cinq à neuf ans et six mois à l'âge de dix ans. Conséquemment, on peut retrouver des différences systématiques entre les estimations démographiques d'âge scolaire (fondées sur l'année scolaire) et celles fondées sur les années révolues. Bien que cela puisse éventuellement fausser le TNS+ de quelques points de pourcentage, on s'attend à ce que l'altération reste minimale puisque la répartition par âge devrait demeurer relativement stable au cours d'une période de six mois.

**Recommandation 12 :** *On devrait procéder à un examen des modalités de définition de l'âge scolaire primaire par les pays et des répercussions de la déclaration de l'âge sur les données d'inscription, les données d'enquête et la définition de la Division de la population des Nations Unies. Au besoin, des études supplémentaires devraient être entreprises afin d'examiner des avenues éventuelles susceptibles de permettre l'ajustement du processus de déclaration ou le redressement des données.*

Une note supplémentaire s'impose sur le moment retenu pour le déroulement de l'EDS : les tableaux 1 et 2 contiennent les dates d'opérations sur le terrain de l'EDS et les dates de début et de fin des années scolaires respectivement. Il y a lieu de noter que dans un certain nombre de pays, l'EDS s'étend de la fin d'une année scolaire au début de la suivante. En outre, la question relative à la fréquentation scolaire est posée rétrospectivement, et elle porte sur l'année qui précède la date de l'enquête, ce qui signifie que la période de l'enquête peut, dans les faits, s'étendre sur deux à trois années scolaires. Il serait préférable que l'EDS soit limitée à une seule année scolaire, ce qui diminuerait la confusion entourant l'âge et la fréquentation. Cependant, étant donné que l'EDS n'a pas été spécifiquement conçue pour servir d'enquête sur l'éducation et qu'il porte davantage sur la santé, il serait surprenant que les pays décident de modifier le déroulement de ces enquêtes de façon à ce qu'elles se conforment à l'année scolaire. Il y a lieu de noter que ce défaut a été corrigé dans le présent rapport par l'utilisation de « taux combinés » pour le TNS+ chaque fois qu'il était comparé avec le TNF+. Nous

recommandons aux analystes qui souhaiteraient comparer l'inscription et la fréquentation d'effectuer le même genre d'ajustement.

Enfin, un mot s'impose sur la façon dont les organismes internationaux déterminent la date de référence des données sur l'âge qui accompagnent celles sur l'inscription. Alors que l'ISU demande aux pays de soumettre l'âge des enfants au moment où débute l'année scolaire, l'OCDE et Eurostat (qui recueillent les données d'inscription pour leurs pays-membres) demandent aux pays de soumettre l'âge des enfants au 31 décembre. Cela peut donner lieu à une grave sous-estimation des TNS. Par exemple, en Allemagne (un pays de l'OCDE), les enfants sont admis à l'école primaire s'ils atteignent l'âge de six ans au 1<sup>er</sup> juillet (pour une année scolaire débutant le 1<sup>er</sup> septembre). Comme la cohorte des enfants âgés de cinq ans est plus vieille de six mois à la date de référence du 31 décembre, près de la moitié des enfants âgés de cinq ans au 1<sup>er</sup> juillet seront enregistrés comme ayant six ans au 31 décembre et seront consignés comme ne s'étant pas inscrits, ce qui occasionnera un sous-dénombrement.

Ce problème de l'âge des enfants est exacerbé par le fait que l'Allemagne se démarque par une durée relativement courte des études primaires (de six ans à neuf ans ou quatre années). En effet, au cours des dernières années, le chiffre du TNS pour l'Allemagne a atteint un creux de 85 %, ce qui ne semble pas très crédible. Pour remédier à ce problème, l'OCDE et Eurostat devraient envisager d'ajuster la date de référence pour la déclaration de l'âge pour qu'elle corresponde au début de l'année scolaire ce qui, conséquemment, la ferait correspondre à la pratique en usage à l'ISU. Étant donné qu'aucun pays de l'OCDE ou de l'UE ne se retrouvait parmi les 16 pays analysés dans ce document, il est possible que cet enjeu ne suscite pas ici de préoccupation immédiate. Toutefois, de façon plus générale, cela demeure source d'inquiétude pour l'ISU puisqu'il dépend des données d'inscription de l'OCDE et d'Eurostat et des données correspondantes par âge.

***Recommandation 13*** : *L'OCDE et Eurostat devraient envisager l'ajustement de la date de référence pour la déclaration de l'âge pour qu'elle corresponde au début de l'année scolaire de façon à ce qu'elle se conforme à la date de référence pour les renseignements sur l'inscription.*

## 7. Conclusions

Dans ce document, les données provenant de 16 pays sur l'inscription et la fréquentation de l'école primaire ont été examinées en tant que fondement d'une enquête relative aux sources éventuelles d'erreurs pouvant contribuer à la présence d'écart entre le TNF+ et le TNS+ et, en dernier ressort, à des estimations inexactes des enfants non scolarisés. Plusieurs des dix études que l'on retrouve ici ont rapproché dans une certaine mesure des divergences qui avaient été observées. C'est au Bangladesh qu'est survenu l'écart le plus considérable. L'étude 9 illustre que les données sur la fréquentation passée n'étaient pas saisies et que cela pouvait contribuer à une sous-déclaration globale portant sur la fréquentation scolaire dans ce pays. C'est pourquoi, en ce qui concerne le Bangladesh, les données administratives sur l'inscription pourraient constituer une source plus crédible. Cependant, l'étude 4 démontre que les révisions de recensement (2002 par opposition à 2004) ont eu un effet considérable au Bangladesh sur les estimations successives d'inscription.

En Indonésie, au Mozambique, en Namibie, au Rwanda et au Vietnam, les différences entre le TNS+ et le TNF+ restent inférieures à 5 % en valeur absolue (tableau 1), mais une décomposition a démontré l'existence d'écart substantiels entre les numérateurs provenant des deux sources de données correspondantes (les données de source administratives par opposition aux données d'enquête du tableau 3) et entre les dénominateurs des deux sources de données correspondantes (les données de recensement par opposition aux données d'enquête du tableau 4). Au Vietnam, il existe de toute évidence un problème avec l'absence d'un facteur multiplicatif pour les poids finaux : tant le numérateur que le dénominateur affichent des différences de près de 50 % entre les deux sources de données correspondantes. Pour les quatre pays restants, l'étude 6 illustre que la stratification *a posteriori* des poids pourrait permettre un réaligement d'un écart éventuel de la représentativité de l'échantillon, dans la mesure où les données d'enquête demeurent la source commune de données pour les numérateurs et les dénominateurs.

En Guinée, l'étude 9 illustre le problème qui peut avoir été occasionné par le moment retenu pour la tenue de l'enquête, lorsqu'elle recoupe les mois des vacances, sous la forme d'une certaine sous-déclaration de la fréquentation actuelle et d'une sur-déclaration de la fréquentation passée. Cela constitue aussi un problème pour le Rwanda.

En Côte d'Ivoire, en Égypte et en Tanzanie, on retrouve des différences négligeables entre les estimations du dénominateur (tableau 4) mais de substantielles divergences entre les estimations du numérateur (tableau 3), ce qui laisse présager de la présence possible d'un problème au niveau des données d'inscription. Cependant, cette situation demeure nébuleuse et il se pourrait que l'on doive approfondir l'examen de cette question. On notera avec intérêt que, bien que des différences négligeables distinguent les estimations du dénominateur (fondées sur les données de recensement par opposition aux données d'enquête), l'étude 3 (tableau 8) démontre curieusement, qu'en ce qui a trait à l'Égypte, l'algorithme d'interpolation de Sprague a produit une variation considérable des estimations fondées sur le recensement (une des deux sources de données). En outre, l'étude 4 illustre l'effet modeste suscité par les révisions de recensement (2002 par rapport à 2004) sur les estimations relatives à l'inscription en Côte d'Ivoire.

Quant aux pays restants considérés dans le cadre de ce rapport (le Burkina Faso, le Ghana, l'Inde, le Kenya, le Mali et le Nigéria), une cohérence raisonnable semble exister

entre les deux sources de données. Dans le cas de ces pays, peu de problèmes perceptibles semblent se présenter au niveau de la qualité des données, à tout le moins en ce qui concerne les études entreprises dans le cadre de ce rapport.

En conclusion, ce document s'efforce d'effectuer une recherche sur certains enjeux importants relatifs à la qualité des données qui surgissent lors de l'utilisation de données d'enquête-ménage par opposition à des données de source administrative (combinées à des données de recensement) pour mesurer la participation scolaire. Bien que ce rapport ne se veuille pas exhaustif, il met en lumière de nombreuses conclusions importantes et il recommande des mesures correctives qui se répercuteront à la fois sur les commanditaires et sur les organismes responsables de la mise en œuvre des enquêtes-ménages (incluant l'UNICEF et Macro International), les responsables secondaires de la cueillette des données administratives (l'ISU) et l'organisme des Nations Unies responsable des données démographiques à l'échelle mondiale (Division de la population des Nations Unies).

Les recommandations formulées dans le rapport ne visent nullement à critiquer le travail des organismes qui y sont mentionnés, et il est entendu que certaines des recommandations nécessiteraient des ressources considérables pour provoquer l'avènement des changements suggérés. Mais ces changements amélioreraient leur cohérence et faciliteraient en général leur utilité.

## Références

Sarndal, C. E., B. Swensson and J. Wretman (1992). *Model Assisted Survey Sampling*. New York : Springer-Verlag.

Singh, A. (2006). "Estimation of Primary Age Children Out-of-School by Combining Cross-Sectional and Longitudinal Data from Administrative and Sample Survey-Based Sources". Internal document. Montréal: UNESCO Institute for Statistics.

Sprague, T. (1881), "Explanation of a new formula for interpolation", *Journal of the Institute of Actuaries*, Vol. 22.

Division de statistique des Nations Unies (DSNU) (2008). "UN Principles and Recommendations for Population and Housing Censuses, Revision 2". New York : Division de statistique des Nations Unies.

UNESCO (2002). *Éducation pour tous: le monde est-il sur la bonne voie? Rapport mondial de suivi sur l'ÉPT*. Paris : UNESCO.

UNESCO-ISU (2005). *Enfants non-scolarisés : Mesure de l'exclusion de l'enseignement primaire*. Montréal : Institut de statistique de l'UNESCO.

UNESCO-ISU (2007). *Recueil de données mondiales sur l'éducation 2007*. Montréal : Institut de statistique de l'UNESCO.

UNICEF (2008). *La situation des enfants dans le monde 2008*. New York : UNICEF.

Division de population des Nations Unies (2002 and 2004). *World Population Prospects (2002 and 2004 Revisions)*. New York : UN Population Division.

Division de population des Nations Unies (2008). *World Population Prospects (2008 Revision)*. New York : UN Population Division.

Wolter, K. M. (1995). *Introduction to Variance Estimation*. New York : Springer-Verlag.

Le nombre d'enfants non scolarisés est une des statistiques les plus utilisées pour faire le suivi des progrès vers l'enseignement primaire universel (EPU) d'ici 2015 : il est donc essentiel d'appliquer une méthodologie méticuleuse lorsque l'on mesure la participation scolaire dans l'enseignement primaire. Traditionnellement, les données d'inscriptions issues des dossiers administratifs sont utilisées pour calculer ces indicateurs. Cependant, depuis une décennie, les statisticiens se tournent vers une autre source de données pour alimenter leurs estimations: les données de fréquentation scolaire disponibles dans les enquêtes-ménages.

Le Document technique N° 4 passe en revue les sources de données et la méthodologie utilisée par l'Institut de statistique de l'UNESCO pour estimer le nombre d'enfants non scolarisés. Ce nombre global ainsi que les sommes régionales qui en sont tirées sont largement utilisés par les gouvernements nationaux et agences des Nations Unies pour suivre les progrès vers l'Éducation pour tous (ÉPT).

Le rapport décrit le taux net de scolarisation ajusté et le taux net de fréquentation ajusté, deux indicateurs de participation scolaire pour lesquels les possibles sources d'erreurs d'usage sont discutées. Il étudie également les écarts entre différentes estimations de population, et leur impact sur la mesure de la participation scolaire. Les auteurs examinent les poids d'enquête, les erreurs d'échantillonnage et de non-échantillonnage des enquêtes-ménages, tout comme les interactions entre le calendrier de la collecte et celui de l'année scolaire. En conclusion, le document compare les divergences dans les données d'âge selon les sources, et leurs effets sur les calculs.



INSTITUT  
*de*  
STATISTIQUE  
*de l'UNESCO*

**Institut de statistique de l'UNESCO**

C.P. 6128, Succursale Centre-Ville

Montréal (Québec) H3C 3J7

Canada

<http://www.uis.unesco.org>

L'Institut de statistique de l'UNESCO (ISU) est le bureau de statistique de l'Organisation des Nations Unies pour l'Éducation, la Science et la Culture (UNESCO) et le dépositaire de l'ONU de statistiques internationales comparables sur l'éducation, la science, la technologie, la culture et la communication.