

Évaluation de l'efficacité des méthodes de traitement des valeurs manquantes dans le contexte d'échelles de mesure de type Likert

Michel Rousseau et Richard Bertrand

Volume 28, numéro 3, 2005

URI : <https://id.erudit.org/iderudit/1087029ar>

DOI : <https://doi.org/10.7202/1087029ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (imprimé)

2368-2000 (numérique)

[Découvrir la revue](#)

Citer cet article

Rousseau, M. & Bertrand, R. (2005). Évaluation de l'efficacité des méthodes de traitement des valeurs manquantes dans le contexte d'échelles de mesure de type Likert. *Mesure et évaluation en éducation*, 28(3), 27–49.
<https://doi.org/10.7202/1087029ar>

Résumé de l'article

La présence de réponses manquantes pour certains items d'une échelle de mesure est un phénomène que tout chercheur est susceptible de rencontrer au cours de ses travaux. Malgré l'ampleur de ce problème, les connaissances quant aux biais ou à l'efficacité des méthodes de traitement sont encore très restreintes. La présente étude vise à faire évoluer les connaissances et les pratiques concernant le traitement des valeurs manquantes dans le contexte d'utilisation d'échelles de type Likert. Ce projet de recherche est basé sur une méthodologie de type Monte Carlo et compare l'efficacité de cinq méthodes de traitement. L'imputation multiple, méthode de traitement qui est en voie de devenir la norme, se montre comme une des méthodes de traitement les plus efficaces, minimisant la présence de biais autant pour ce qui est du coefficient alpha de Cronbach que du score à l'échelle.

Évaluation de l'efficacité des méthodes de traitement des valeurs manquantes dans le contexte d'échelles de mesure de type Likert

Michel Rousseau

Université de Moncton

Richard Bertrand

Université Laval

MOTS CLÉS: Valeurs manquantes, théorie classique des tests, méthodes de traitement des valeurs manquantes, imputation multiple

La présence de réponses manquantes pour certains items d'une échelle de mesure est un phénomène que tout chercheur est susceptible de rencontrer au cours de ses travaux. Malgré l'ampleur de ce problème, les connaissances quant aux biais ou à l'efficacité des méthodes de traitement sont encore très restreintes. La présente étude vise à faire évoluer les connaissances et les pratiques concernant le traitement des valeurs manquantes dans le contexte d'utilisation d'échelles de type Likert. Ce projet de recherche est basé sur une méthodologie de type Monte Carlo et compare l'efficacité de cinq méthodes de traitement. L'imputation multiple, méthode de traitement qui est en voie de devenir la norme, se montre comme une des méthodes de traitement les plus efficaces, minimisant la présence de biais autant pour ce qui est du coefficient alpha de Cronbach que du score à l'échelle.

KEY WORDS: Missing values, classical test theory, missing values treatment, multiple imputation

The presence of non-response for some items on a scale of measurement is a phenomenon which any researcher is suitable to meet during his work. Despite the extent of this problem, knowledge about bias or effectiveness of the missing data treatment is still very restricted. The aim of the present study is to evolve knowledge and practices concerning the treatment of the missing values in the context of Likert-type scales. This research project, based on a Monte Carlo methodology, compares the effectiveness of five methods of treatment. The multiple imputation, the method which is in the process of becoming the standard, is shown as being one of the most effective methods of treatment, minimizing the presence of bias as much on the level of the coefficient alpha of Cronbach that on the scale score.

PALAVRAS-CHAVE: Não-respostas (missing values), teoria clássica dos testes, métodos de tratamento dos missing values, imputação múltipla

A presença de não respostas (missing values), para certos itens de uma escala de medida é um fenómeno a que está sujeito qualquer investigador, ao longo dos seus trabalhos. Apesar da amplitude deste problema, os conhecimentos relativos aos enviesamentos e à eficácia dos métodos de tratamento são ainda muito restritos. O presente estudo visa fazer evoluir os conhecimentos e as práticas relativos ao tratamento das não respostas no contexto de utilização de escalas de tipo likert. Este projecto de investigação assenta numa metodologia de tipo Monte Carlo e compara a eficácia de cinco métodos de tratamento. A imputação múltipla, método de tratamento que está em vias de vir a ser a norma, revela-se como um dos métodos de tratamento mais eficazes, minimizando os enviesamentos tanto ao nível do coeficiente alpha de Cronbach, quanto ao do score da escala.

Introduction

Le but de toute étude scientifique est de produire des connaissances à partir des résultats de recherche qui soient le plus possible libres d'erreurs ou de biais. Une pléthore de sources d'erreurs ont été dénombrées au fil des ans. Parmi celles-ci se retrouvent les erreurs relatives à la validité interne d'un projet de recherche, les erreurs liées à l'échantillonnage ainsi que les erreurs de mesure. Les chercheurs spécialisés en méthodologie se sont efforcés de réduire ces sources d'erreurs et de grandes améliorations ont été proposées. Toutefois, une source de biais encore peu explorée demeure celle associée aux erreurs causées par la présence de données manquantes. Cela est d'autant plus vrai dans le cas des échelles de mesure de type Likert (McDonald, Thurston & Nelson, 2000). Enders (2004) rapporte avoir relevé uniquement trois études traitant du problème des valeurs manquantes dans le cas d'échelles de mesure de type Likert. Le très peu d'intérêt porté à ce problème est d'autant plus étonnant que les échelles de mesure revêtent une importance particulière dans la recherche scientifique (Enders, 2004) et que les biais importants que peut causer la présence de valeurs manquantes ont été cernés depuis près de 30 ans par Rubin (1976). Comme le soulignent Schaffer et Graham (2002), il est d'autant plus pertinent de tenir ce type d'étude qu'une méthode de traitement des valeurs manquantes doit être testée dans le contexte spécifique des analyses qui sont visées. Ainsi, une méthode de traitement peut être efficace pour des items de type dichotomique, mais se révéler inefficace pour dans le cas d'échelles de mesure de type Likert.

Note des auteurs : Toute correspondance peut être adressée par courriel à l'adresse suivante : [roussem@UMoncton.CA]

Contexte théorique

Une donnée est considérée comme étant manquante lorsque, en inspectant une matrice de données, on constate qu'une valeur n'est pas observée alors qu'il aurait été logique d'en observer une (Little & Rubin, 2002). La présence de données manquantes est un phénomène qui est susceptible d'être observé dans pratiquement toute recherche entreprise dans le domaine de l'éducation ou des sciences humaines. Bien que cette situation soit inévitable, l'importance des biais causés par ce type de données est encore peu explorée.

Les données manquantes proviennent de sources très variées. Ces données peuvent être classées en trois grandes catégories, soit les sujets manquants, les occasions manquantes et les valeurs manquantes (Van der Kamp & Bijleveld, 1998). Les sujets manquants sont définis comme étant ceux qui ont été sélectionnés pour participer à une étude, mais qui n'y participent pas pour différentes raisons. Ce type de données manquantes regroupe, par exemple, les sujets qui refusent de participer à une étude ou de répondre à un sondage, les sujets avec lesquels il était impossible de prendre contact au moment du recrutement et ceux qui n'étaient pas disponibles ou qui n'étaient pas en mesure de participer. La présence de sujets manquants affecte directement la représentativité de l'échantillon par rapport à la population cible, limitant d'autant la capacité de généralisation des résultats.

Les occasions manquantes constituent un type de données manquantes qui peuvent être observées lorsque le devis de recherche inclut plus d'un temps de mesure. Dans cette situation, les données sont disponibles pour les sujets à un temps de mesure, tandis qu'elles sont complètement manquantes à un autre temps de mesure. Les occasions manquantes peuvent être causées par des sujets qui déménagent, qui meurent, qui refusent de continuer à participer à l'étude ou pour d'autres raisons. Tout comme pour les sujets manquants, ce type de données manquantes affectera la représentativité de l'échantillon.

Finalement, bien que les sujets aient accepté de participer à une étude, il est toujours possible que ceux-ci présentent des données manquantes. En effet, il est courant que des réponses à certaines questions ou items d'une échelle de mesure soient manquantes pour certains individus, causant ce qu'il est convenu d'appeler des valeurs manquantes. Celles-ci peuvent tout aussi bien émaner de sujets qui refusent de répondre à certaines questions que de sujets qui oublient simplement d'y répondre. La présente étude s'intéresse plus particulièrement à ce dernier type de données manquantes.

Les valeurs manquantes posent problème, tout particulièrement lors de l'analyse des résultats. De manière générale, l'information doit être disponible pour toutes les variables pour qu'un sujet soit inclus dans les analyses statistiques. La plupart des méthodes statistiques ont été élaborées de manière à analyser une matrice de données rectangulaire dont aucune donnée n'est manquante (Little & Rubin, 2002).

Cette dernière particularité illustre bien le problème que pose la présence de valeurs manquantes dans une banque de données. Les modèles de mesure utilisés pour analyser les données d'échelle de type Likert n'échappent pas à cette difficulté. Peu importe le modèle utilisé, l'estimation de la mesure du trait psychologique et des qualités psychométriques des items peut être effectuée uniquement pour les individus qui ont fourni des réponses à tous les items de l'échelle (Huisman, 2000). Le coefficient alpha de Cronbach, à titre d'exemple, est calculé uniquement pour les sujets ayant répondu à l'ensemble des items qui composent l'échelle. Le coefficient ainsi obtenu n'est donc valide que pour la proportion d'individus présentant des patrons de réponses complets.

Il est reconnu que le mécanisme causant la présence de données manquantes est le facteur ayant le plus d'influence sur l'importance des biais causés par les valeurs manquantes (Little & Rubin, 2002). Rubin (1976) a défini trois mécanismes pouvant expliquer la présence de données manquantes dans une banque de données. Les données sont dites manquantes selon un mécanisme complètement aléatoire (*missing completely at random*) lorsque la probabilité qu'une valeur soit manquante sur une variable particulière J est indépendante de la valeur de J elle-même ou de la valeur de n'importe quelle autre variable incluse dans la matrice de données. Ainsi, les valeurs manquantes ne sont reliées à aucune valeur des variables contenues dans la base de données. Les données sont dites manquantes selon un mécanisme aléatoire (*missing at random*) lorsque la probabilité qu'une valeur soit manquante sur une variable particulière J est indépendante de cette variable J mais dépendante d'une autre variable comprise dans la base de données. Enfin, elles sont dites manquantes selon un mécanisme non aléatoire (*not missing at random*) lorsque la probabilité d'observer une valeur manquante sur la variable J est dépendante de la valeur de J elle-même. Pour ce dernier mécanisme, il s'agit de la situation où les biais sont les plus importants (Little & Rubin, 2002).

Bien que souhaitable, il n'est malheureusement pas possible de définir le véritable mécanisme qui sous-tend la présence d'une valeur manquante (Little & Rubin, 2002; Schaffer, 1998). Pour être en mesure d'effectuer une telle démonstration, il serait essentiel d'obtenir de l'information sur les données manquantes afin de tester si celles-ci ne sont pas dépendantes de leur réelle valeur ou de celle de toute autre variable. Or, si tel était le cas, le problème des données manquantes serait *ipso facto* réglé. En conséquence, les trois mécanismes sont considérés comme étant des postulats. Il est reconnu que les données manquantes sont rarement manquantes selon un mécanisme complètement aléatoire (Little & Rubin, 2002) et qu'elles respectent très peu souvent les conditions des données manquantes selon un mécanisme aléatoire (Allison, 2001).

Diverses méthodes de traitement des données manquantes ont été suggérées pour tenter de faire face à ce problème. Leur efficacité rapportée par diverses études semble être très variable d'une méthode à une autre. L'efficacité d'une méthode de traitement des données manquantes est définie comme étant sa capacité à fournir des estimés de paramètres ainsi que des erreurs types qui soient le moins biaisés possible. Tel que soulevé par Allison (2001), une méthode de traitement ne peut jamais éliminer complètement les biais relatifs à la présence de données manquantes : au mieux permet-elle de minimiser l'impact de ces biais.

Une des méthodes de traitement qui est sûrement la plus populaire consiste à retirer des analyses tout individu présentant une valeur manquante sur au moins une variable contenue dans le modèle d'analyse. Cette méthode de traitement, appelée analyse des cas complets, peut être effectuée à partir de deux techniques spécifiques que nous appellerons l'élimination par liste (*listwise deletion*) et l'élimination par paire (*pairwise deletion*). L'élimination par liste, par sa simplicité et par le fait qu'elle constitue l'option par défaut de bien des logiciels, se présente comme étant la méthode de traitement la plus répandue.

D'autres méthodes de traitement semblent aussi être très populaires pour traiter les valeurs manquantes dans une échelle de type Likert. L'imputation par la moyenne du sujet, par exemple, propose de remplacer la ou les valeurs manquantes sur un ou plusieurs items pour un individu à partir de la moyenne des autres items auxquels a répondu cet individu. L'imputation par la moyenne de l'item, quant à lui, suggère de remplacer les valeurs manquantes sur un item à partir de la moyenne des réponses observées pour ce même item.

Les premières études portant sur l'évaluation des méthodes de traitement des valeurs manquantes utilisaient principalement les méthodes précitées, dites classiques. Toutefois, au cours des dernières années, l'intérêt envers les méthodes «modernes» de traitement des valeurs manquantes s'est accru, en particulier envers les méthodes basées sur l'estimation par maximum de vraisemblance ainsi qu'envers les méthodes d'imputation multiple (Enders, 2003). Des travaux tels que ceux de Sinharay, Stern et Russel (2001), de Fichman et Cummings (2003) et de Schaffer et Graham (2002) ont permis de montrer la supériorité de l'imputation multiple sur des méthodes de traitement plus classiques. Cette dernière méthode est caractérisée par l'obtention de plus d'une matrice de données complètes et plausibles, permettant de refléter l'incertitude reliée au fait que les données complètes proviennent d'un mécanisme d'imputation.

Downey et King (1998) semblent être les premiers auteurs à s'être intéressés à cette problématique. Leur étude comparait l'efficacité des méthodes d'imputation par la moyenne du sujet et par la moyenne de l'item. Ils concluent que ces deux méthodes sont tout aussi efficaces lorsque la proportion de valeurs manquantes est inférieure à 20%, mais que la méthode d'imputation par la moyenne de l'item est plus efficace si la proportion de valeurs manquantes est supérieure à 20%. Cette conclusion est à l'inverse de celles tirées des autres études qui ont été effectuées. La majorité des études démontrent que cette dernière méthode est la pire méthode de traitement et serait même à proscrire (Enders 2004; Huisman, 2000; Roth, Switzer & Switzer, 1999). La méthode d'imputation par la moyenne du sujet, quant à elle, semble plus prometteuse. En effet, en comparaison avec les autres méthodes de traitement évaluées dans les différentes études, cette méthode de traitement serait une de celles présentant le moins de biais (Enders, 2003; Huisman, 2000; McDonald, Thurston & Nelson, 2000; Roth, Switzer & Switzer, 1999). Raaijmakers (1999), quant à lui, propose une nouvelle méthode pour traiter les valeurs manquantes. Cette méthode, appelée imputation par la moyenne relative (*Relative mean substitution*), aurait démontré une efficacité plus élevée que des méthodes classiques telles que la méthode d'imputation par la moyenne du sujet.

Selon les articles consultés, les méthodes modernes de traitement n'ont été évaluées que par Enders (2003, 2004). Dans ses deux études, cet auteur conclut que la méthode *expectation-maximization* (EM) est plus efficace si les données manquantes sont manquantes selon un mécanisme complètement aléatoire ou selon un mécanisme aléatoire, mais que la méthode d'imputation par la moyenne du sujet demeure plus efficace dans le cas où les données sont manquantes selon un mécanisme non aléatoire.

Certaines limites relatives aux études impliquant les méthodes de traitement des données manquantes doivent être soulignées. Premièrement, quelques-unes de ces études, c'est le cas notamment de celles de McDonald, Thurston et Nelson (2000) et de Downey et King (1998), utilisent une méthode de simulation des données et vont effacer de manière complètement aléatoire certaines valeurs pour simuler des matrices des données incomplètes. Tandis que d'autres études, en plus d'utiliser un mécanisme aléatoire pour effacer les données, simulent une situation où les données sont manquantes selon un mécanisme aléatoire (Enders, 2004; Roth, Switzer & Switzer, 1999). Étant donné qu'il est fort probable que les données soient manquantes selon un mécanisme non aléatoire (Little & Rubin, 2002), la généralisation des résultats tirés de ces études semble plutôt limitée. De plus, en voulant contrôler la valeur des corrélations inter-items, les chercheurs ont fixé leurs valeurs de manière à ce qu'elles soient identiques pour chaque paire d'items de l'échelle (Enders, 2003, 2004; Roth, Switzer & Switzer, 1999). Cette situation est hautement hypothétique et peut avoir une influence sur l'efficacité des méthodes utilisées. En effet, en fixant la valeur des corrélations inter-items à une valeur identique pour les items d'une échelle, cela revient à simuler une situation où les items sont des indicateurs de même force du trait latent. Or, il semble que dans une étude fondée sur des simulations, il importe de faire varier les corrélations inter-items pour être en mesure de généraliser plus facilement les résultats obtenus à partir de données simulées à des données de recherche réelles.

Outre le mécanisme causant la présence des valeurs manquantes, certains autres facteurs pouvant avoir une influence sur l'efficacité des méthodes de traitement ont été considérés. C'est le cas notamment de la taille de la banque de données, du nombre d'items de l'échelle ainsi que de la proportion des valeurs manquantes. Alors que certaines études soulignent que la valeur du biais diminue lorsque le nombre d'items augmente (Huisman, 2000; McDonald, Thurston & Nelson, 2000; Roth, Switzer & Switzer, 1999), certaines autres études concluent que le nombre d'items n'a pas d'impact sur l'efficacité des méthodes de traitement (Enders, 2003; Raaijmakers, 1999).

Concernant la taille de la banque de données, les études sont unanimes à conclure que ce facteur influencerait très peu les méthodes de traitement. Enfin, selon Huisman (2000), la proportion de valeurs manquantes aurait un impact sur l'efficacité des méthodes de traitement. Plus cette proportion serait faible et plus l'efficacité des différentes méthodes de traitement serait

identique. Il faut toutefois souligner que Huisman fixe cette proportion de valeurs manquantes à 5%, 12% et 20%. Comme il est attendu que lorsque la proportion de valeurs manquantes est faible (près de 10% ou moins) les différences entre les méthodes diminuent, l'effet observé de la proportion de valeurs manquantes pourrait être attribué aux caractéristiques mêmes de son design expérimental.

Objectifs de l'étude

Étant donné la précarité des connaissances concernant l'impact des méthodes de traitement de valeurs manquantes sur les qualités psychométriques d'échelles de mesure de type Likert, outre que les résultats des études déjà réalisées sur le sujet demeurent souvent ambigus et que la plupart de ces études présentent des limites méthodologiques importantes, la tenue d'une évaluation plus systématique de l'efficacité des méthodes de traitement des valeurs manquantes s'est imposée par elle-même. La présente étude a comme objectif principal d'évaluer cinq méthodes de traitement des valeurs manquantes afin de déterminer la ou les méthodes les plus efficaces pour ce qui est du biais des estimés du coefficient alpha de Cronbach et du score à l'échelle. De plus, cette étude vise à évaluer l'influence du nombre d'items, du nombre de sujets et de la proportion de données manquantes sur l'efficacité de ces méthodes de traitement. Puisque, pour conduire une telle étude, il est essentiel de connaître les valeurs réelles des paramètres de la population, le présent projet de recherche est basé sur des données simulées.

Méthode

Méthodes de traitement des valeurs manquantes évaluées

Un grand nombre de méthodes de traitement des valeurs manquantes ont été explorées dans les diverses études relevées dans la documentation. Il est impensable de tenter d'inclure l'ensemble de ces méthodes dans une seule étude. La présente recherche s'est donc intéressée aux méthodes de traitement les plus souvent comparées dans les principales études déjà réalisées. Cette étude vise donc à comparer les cinq méthodes suivantes :

Élimination par liste (EL). Les sujets présentant au moins une valeur manquante sur un des items de l'échelle ont été retirés pour effectuer le calcul du coefficient alpha de Cronbach. Puisque cette méthode constitue bien

souvent celle qui est proposée par défaut dans plusieurs logiciels, elle reflète en réalité la situation où le chercheur n'a pas vraiment conscience d'appliquer une méthode de traitement de valeurs manquantes en bonne et due forme.

L'imputation par la moyenne à l'item (IMI). Les données manquantes sur un item ont été remplacées par la moyenne des réponses à cet item, et ce, pour tous les items de l'échelle.

L'imputation par la moyenne du sujet (IMS). Les données manquantes pour un sujet ont été remplacées par la moyenne des réponses de ce sujet aux autres items, et ce, pour tous les sujets de la banque de données.

L'imputation par équation de régression en tenant compte de l'erreur résiduelle (REG). Les données manquantes pour un item ont été remplacées par des valeurs prédites à l'aide d'une équation de régression qui inclut tous les autres items comme variables prédictives, et ce, pour chacun des items de l'échelle.

L'imputation multiple (IM). Au lieu de créer une seule matrice de données complètes, comme dans le cas de la méthode précédente, cinq matrices de données imputées sont créées. Le choix de cinq matrices constitue l'option qui permet de régler le plus efficacement possible la majorité des situations rencontrées (Schaffer & Graham, 2002). L'imputation des matrices de données est basée sur l'utilisation de la méthode EM.

La méthode EM est une procédure d'estimation itérative en deux étapes. Le processus débute par une estimation initiale de la matrice de covariance. Cette matrice de covariance permet de procéder à l'imputation des valeurs manquantes dans la base de données, ce qui constitue la phase E (*expectation*). À partir de ces données, une nouvelle matrice de covariance est par la suite obtenue, ce qui constitue la phase M (*maximization*). Cette nouvelle matrice de covariance permet de procéder à une nouvelle phase E, et ainsi de suite jusqu'à ce que la différence entre deux matrices de covariance obtenues lors de deux phases M consécutives soit inférieure à un certain critère de convergence. C'est à partir de cette matrice de covariance que seront créées les cinq matrices de données complètes imputées.

Les méthodes d'imputation ont été réalisées à l'aide du logiciel SAS. Plus particulièrement, les méthodes de régression et d'imputation multiple ont été effectuées à partir de la procédure PROC MI, qui a été implantée dans la version 8 du logiciel. Chaque valeur imputée a été arrondie à la valeur plausible la plus près afin de conserver la nature ordinale des items d'échelles de type Likert.

Design expérimental

En plus des méthodes de traitement des valeurs manquantes, l'impact de trois variables indépendantes a été examiné : la taille de l'échantillon (400 sujets et 4 000 sujets), le nombre d'items de l'échelle (quatre items et neuf items) ainsi que la proportion de valeurs manquantes (15 % et 25 %). La taille de l'échantillon a été sélectionnée pour permettre de généraliser les résultats de la présente étude autant à des études faites sur un échantillon de plus petite taille qu'à des études à grande échelle. Pour ce qui est du nombre d'items, comme les paramètres de générations des données pour ces échelles sont obtenus à partir d'un échantillon de très grande taille, soit l'échantillon du premier cycle de l'étude PISA (2000), les nombres d'items possibles étaient contraints par les échelles existant dans cette étude. Dans le cycle 2000 du PISA, le nombre d'items par échelle de type Likert varie entre trois et neuf items. Enfin, en ce qui concerne la proportion de valeurs manquantes, il était important de ne pas fixer cette proportion à une trop petite valeur. Il semble en effet que cette proportion ne doive pas se situer en dessous de 10%, seuil sous lequel les résultats obtenus à partir de l'application de diverses méthodes de traitement des données manquantes tendent à être identiques (Schaffer & Graham, 2002). De plus, pour refléter des situations de recherche réalistes, il importait de ne pas utiliser une proportion de trop grande taille.

Variables dépendantes

Avec ou sans donnée manquante, le but de toute procédure statistique est d'effectuer une inférence valide à propos de la population d'intérêt (Schaffer & Graham, 2000). Ainsi, les analyses ne s'attarderont pas à vérifier si les méthodes de traitement comparées permettent de prédire efficacement les valeurs manquantes aux items, mais concerneront plutôt les estimés du trait latent et les qualités psychométriques des échelles de mesure.

Puisque les qualités psychométriques d'échelles de mesure de type Likert peuvent en théorie être évaluées à partir d'un certain nombre de critères, deux variables dépendantes serviront à cerner l'efficacité des méthodes de traitement. Sur le plan de la fidélité de l'échelle, le choix s'est porté sur l'indice de fidélité estimé à partir du coefficient alpha de Cronbach. De plus, l'utilité première des échelles de mesure étant de fournir un estimé d'un trait psychologique qui sera utilisé pour effectuer les analyses statistiques désirées, une méthode de traitement efficace devrait de ce fait être en mesure de fournir un estimé du trait qui soit le plus près possible de ce qui aurait été observé si les données avaient été complètes. Ainsi, le jugement de l'efficacité des

méthodes de traitement porte sur la capacité des diverses méthodes de traitement à produire, pour l'ensemble des individus de l'échantillon, un estimé du trait psychologique, dont la valeur s'approche de la valeur qui aurait dû être observée. Cet estimé du trait latent est obtenu à partir du score des individus à l'échelle de mesure.

Le coefficient alpha de Cronbach. Les critères de base pour évaluer une procédure d'analyse statistique ont été établis par Neyman et Pearson (1933) et Neyman (1937). Considérons la valeur comme étant le paramètre d'intérêt d'une population et la valeur comme étant son estimé, basé sur un échantillon aléatoire de cette même population. Il est attendu que toute procédure d'analyse statistique adéquate permet à la valeur de se rapprocher de celle de , et ce, autant pour un échantillon particulier que pour la moyenne des calculée à partir d'un grand nombre d'échantillons. Il est souhaitable que le biais, c'est-à-dire la différence entre la moyenne des et la valeur véritable de soit le plus petit possible. Plus cette valeur s'approchera de 0, et plus la méthode de traitement sera jugée efficace.

Le score à l'échelle. La corrélation de Pearson, entre le score obtenu, pour l'ensemble des sujets, à la suite de l'application de chaque méthode de traitement et le score pour la banque de données complètes, est utilisée pour porter un jugement sur l'efficacité des méthodes de traitement évaluées dans cette étude. Étant donné que l'application de la méthode d'élimination par liste ne remplace pas les valeurs qui sont manquantes pour le score classique, au lieu de considérer la somme de tous les items de l'échelle, ce score est calculé à partir de la moyenne des items auxquels chaque individu a répondu. Pour les autres méthodes, ce score moyen sera calculé sur tous les items de l'échelle de mesure à partir des données complètes et des valeurs imputées.

Procédure de génération et d'effacement des données

La simulation des données, basée sur celle effectuée par Enders (2004), s'est effectuée en deux étapes. Cette simulation a été réalisée à l'aide du logiciel SAS. Une banque de données a été créée à partir des sujets ayant fourni une réponse à tous les items pour les deux échelles ciblées dans l'enquête PISA 2000, soit une échelle de quatre items et une échelle de neuf items. Cette banque de données est constituée de 211 818 sujets. Aux fins de l'étude, cette banque est alors considérée comme étant la population pour laquelle nous désirons obtenir des estimés à partir des échantillons qui seront tirés. Dans la première étape de la simulation, des échantillons de 400 et 4 000 sujets ont été tirés aléatoirement avec remise à partir de la banque de 211 818 individus, et

ce, de manière indépendante pour les échelles de neuf et de quatre items. Ainsi, un premier devis expérimental de quatre cellules, soit deux tailles d'échantillon combinées à deux tailles d'échelle de mesure, est obtenu. Pour chacune de ces cellules, 100 échantillons ont été tirés.

Dans un deuxième temps, à partir de ces échantillons, une portion des données a été effacée des banques pour simuler les valeurs manquantes. Cette opération a été répétée à deux reprises en fonction de la proportion des données qui doivent être enlevées, soit 15% et 25%. Les données manquantes ont été déterminées à partir d'un seul mécanisme qui simule une condition de données manquante selon un mécanisme non aléatoire.

Pour l'échelle de quatre items, les données manquantes ont été imposées uniquement à deux items de l'échelle tandis que pour celle de neuf items, elles ont été imposées à cinq items de l'échelle: ces items sont appelés items-cible. Pour chacune des deux échelles, les items présentant la moyenne la plus faible ont été utilisés comme item-cible. Selon Enders (2004), il est plus vraisemblable que les données manquantes soient observées sur une portion précise des items d'une échelle plutôt que sur l'ensemble des items de l'échelle. Il est à préciser que la proportion de données qui doivent être effacées (15% et 25%) représente la proportion de valeurs manquantes pour chacun des items-cible plutôt que la proportion de valeurs manquantes pour l'ensemble des données.

Pour simuler une condition de données manquantes selon un mécanisme non aléatoire, la probabilité d'observer une valeur manquante pour un item-cible est définie comme étant dépendante du trait latent mesuré par l'échelle. Ainsi, selon ce mécanisme, la probabilité d'observer une valeur manquante sur un item est dépendante de la valeur du trait d'un individu ainsi que des caractéristiques de l'item. Pour procéder à l'effacement des valeurs, une probabilité a été assignée à chaque sujet de la banque de données. Les scores ont par la suite été ordonnés du plus petit au plus grand. Pour chaque item-cible, la procédure d'effacement des données débute par le premier sujet de la banque ainsi ordonnée: si la probabilité observée pour ce sujet est supérieure à une probabilité fixée arbitrairement pour procéder à l'effacement des données, la réponse pour ce sujet est remplacée par une valeur manquante. La procédure continue en descendant dans le fichier de données tant que la proportion désirée de valeurs manquantes (15% ou 25%) n'est pas atteinte.

Analyses statistiques des données

Pour analyser l'impact des variables indépendantes sur le biais en valeur absolue et sur le recouvrement des scores d'échelles, des analyses de la variance (ANOVA) à trois facteurs ont été utilisées. Les variables indépendantes de cette analyse sont la méthode de traitement des valeurs manquantes (MTVM), le nombre de sujets (NS) ainsi que la proportion de valeurs manquantes (PVM). Les analyses sont effectuées de manière indépendante pour l'échelle de quatre items et celle de neuf items.

Le fait de travailler avec une très grande banque de données, soit 100 sujets par cellule ou un total de 2 000 sujets pour chaque ANOVA, confère certains avantages. Le respect des conditions d'application de ces méthodes d'analyse devient une préoccupation moins importante. En effet, même si la variable dépendante n'est pas distribuée selon la loi normale pour chacune des populations définies par le devis d'analyse, les résultats risqueront tout de même d'être valides dans la mesure où l'échantillon est de grande taille (Green & Salkin, 2003). De plus, comme il s'agit d'un devis équilibré, un manque d'homogénéité de la variance de la variable dépendante pour chaque cellule du devis n'aura pas de conséquences sérieuses (Green & Salkin, 2003). Toutefois, cette grande taille constitue aussi une limite. Les tests d'hypothèses devenant trop puissants, toute différence, même quasiment nulle sur le plan pratique, risque d'être vue comme une différence statistiquement significative. Ainsi, au lieu de s'attarder à l'interprétation des valeurs du p , la taille de l'effet (η^2) est utilisée pour découvrir les facteurs présentant les différences les plus importantes. La taille de l'effet est un indice standardisé qui peut prendre des valeurs se situant entre 0 et 1. Pour un facteur donné, plus la valeur de la taille de l'effet s'approche de 1 et plus l'effet dû à ce facteur est important.

Résultats

Le coefficient alpha de Cronbach

Les résultats de l'ANOVA pour les deux échelles de mesure sont présentés au tableau 1. Il est observé, pour l'échelle de quatre items, que le facteur de la méthode de traitement des valeurs manquantes (MTVM) présente une valeur de taille d'effet relativement plus grande que celles observées pour tous les autres effets pour cette ANOVA. En effet, la valeur de la taille de l'effet pour la méthode de traitement est de,827 comparativement à l'effet d'interaction entre la méthode de traitement et la proportion de valeurs manquantes (MTVM*PVM) qui présente une valeur de,038.

Tableau 1
**Valeurs des tailles de l'effet pour le coefficient alpha de Cronbach
des échelles de quatre items et de neuf items**

	MTVM	NS	PVM	MTVM *	MTVM *	NS *	MTVM *
				NS	PVM	PVM	NS *
							PVM
4 items	,827	,005	,032	,020	,038	,000	,000
9 items	,956	,001	,134	,005	,501	,000	,000

MTVM = méthode de traitement des valeurs manquantes, NS = nombre de sujets, PVM = proportion de valeurs manquantes

Ces résultats suggèrent donc que seules les différences entre les méthodes de traitement se doivent d'être considérées. Les résultats présentés au tableau 2 indiquent que la méthode d'élimination par liste se distingue nettement des autres méthodes. En effet, la valeur moyenne du biais est beaucoup plus élevée que pour les autres méthodes.

Tableau 2
**Valeur moyenne et écart type du biais du coefficient alpha de Cronbach
pour l'échelle de quatre items**

	400 sujets		4 000 sujets	
	15%	25%	15%	25%
EL	-,303 (0,08)	-,325 (0,09)	-,300 (0,03)	-,316 (0,03)
IMS	-,022 (0,05)	,015 (0,05)	-,020 (0,02)	,019 (0,01)
IMI	-,197 (0,06)	-,181 (0,06)	-,194 (0,02)	-,174 (0,02)
REG	-,097 (0,06)	,069 (0,07)	-,083 (0,02)	-,053 (0,03)
IM	-,105 (0,06)	-,084 (0,06)	-,099 (0,02)	-,078 (0,02)

En ce qui concerne l'échelle de neuf items (tableau 1), la valeur de la taille de l'effet d'interaction MTVM*PVM est relativement élevée (,501). La valeur de la taille de l'effet simple de la méthode de traitement est toutefois beaucoup plus élevée, avec,956. Étant donné la présence d'un effet d'interaction qui peut être important, il importe d'analyser tout d'abord cet effet en détail. Les résultats présentés au tableau 3 permettent de procéder à l'interprétation de cet effet.

Il semble exister une différence plus importante entre la valeur moyenne du biais en fonction de la proportion de valeurs manquantes dans le cas de la méthode d'élimination par liste que pour les autres méthodes de traitement. Par exemple, pour la méthode d'élimination par liste, dans la situation où la banque de données compte 400 individus, la valeur moyenne du biais est de $-,121$ lorsque la proportion de valeurs manquantes est de 15% et de $-,184$ lorsque cette proportion est de 25%. En comparaison, la valeur moyenne du biais passe de $,010$ à $,018$ pour la méthode d'imputation par la moyenne du sujet.

Tel que le démontre le tableau 2, pour l'échelle de quatre items, la méthode d'imputation par la moyenne du sujet présente les valeurs moyennes de biais les plus petites par rapport aux autres méthodes de traitement, et ce, peu importe le nombre de sujets et la proportion de valeurs manquantes. En ce qui concerne l'échelle de neuf items, ce sont les méthodes d'imputation par la moyenne du sujet, de régression et d'imputation multiple qui présentent les valeurs moyennes les plus petites. Parmi ces trois méthodes, l'imputation multiple obtient les valeurs les plus petites, et ce, peu importe le nombre de sujets et la proportion de valeurs manquantes. Il est toutefois à noter que les différences entre ces trois méthodes se situent généralement à la troisième décimale.

Tableau 3
Valeur moyenne et écart-type du biais du coefficient alpha de Cronbach pour l'échelle de neuf items

	400 sujets		4 000 sujets	
	15%	25%	15%	25%
EL	$-,121 (0,02)$	$-,184 (0,04)$	$-,117 (0,01)$	$-,179 (0,01)$
IMS	$,010 (0,01)$	$,018 (0,01)$	$,009 (0,003)$	$,018 (0,003)$
IMI	$-,074 (0,01)$	$-,076 (0,01)$	$-,075 (0,01)$	$-,077 (0,01)$
REG	$-,009 (0,01)$	$-,007 (0,01)$	$-,008 (0,004)$	$-,007 (0,01)$
IM	$-,004 (0,01)$	$-,001 (0,01)$	$-,004 (0,003)$	$,001 (0,003)$

Le score à l'échelle

Les résultats de l'analyse des corrélations moyennes, présentés au tableau 4, indiquent que l'effet de la méthode de traitement présente les valeurs de tailles de l'effet les plus élevées. Ces valeurs sont de $,899$ dans le cas de l'échelle de quatre items et de $,870$ pour l'échelle de neuf items. La proportion de valeurs manquantes présente aussi des valeurs de tailles de l'effet assez élevées, soit des valeurs de $,756$ et de $,644$. Enfin, la valeur de l'effet d'interaction MTVM* PVM suggère la présence d'un effet d'interaction qui doit être pris en compte.

Tableau 4
**Valeurs des tailles de l'effet pour les corrélations aux scores classiques
à l'échelle de quatre items et de neuf items**

	<i>MTVM</i>	<i>NS</i>	<i>PVM</i>	<i>MTVM *</i> <i>NS</i>	<i>MTVM *</i> <i>PVM</i>	<i>NS *</i> <i>PVM</i>	<i>MTVM *</i> <i>NS *</i> <i>PVM</i>
4 items	,899	,004	,756	,002	,342	,000	,001
9 items	,870	,001	,644	,006	,478	,000	,002

MTVM = méthode de traitement des valeurs manquantes, NS = nombre de sujets,
PVM = proportion de valeurs manquantes

Le tableau 5 permet d'interpréter cet effet d'interaction. Les résultats indiquent que, bien que les corrélations moyennes soient toujours moins élevées lorsque la proportion de valeurs manquantes est de 25 %, la différence entre les corrélations moyennes calculées pour les deux pourcentages est plus grande dans le cas de la méthode d'imputation par la moyenne du sujet. De plus, il est à noter que cette dernière méthode présente des corrélations moyennes généralement moins élevées que ce qui est observé pour les autres méthodes de traitement. Finalement, il faut remarquer que les corrélations moyennes sont systématiquement inférieures pour l'échelle de quatre items.

Tableau 5
**Valeurs moyennes des corrélations entre les scores classiques
obtenus pour chaque méthode et le score classique obtenu
avec les données complètes**

	<i>Score classique</i>			
	<i>4 items</i>		<i>9 items</i>	
	<i>15%</i>	<i>25%</i>	<i>15%</i>	<i>25%</i>
EL	,976 (0,004)	,955 (0,006)	,989 (0,001)	,983 (0,002)
IMS	,921 (0,011)	,878 (0,012)	,974 (0,005)	,956 (0,006)
IMI	,957 (0,004)	,946 (0,005)	,966 (0,003)	,966 (0,003)
REG	,953 (0,006)	,925 (0,008)	,983 (0,002)	,974 (0,003)
IM	,953 (0,006)	,925 (0,009)	,984 (0,002)	,974 (0,003)

Discussion

Le premier objectif de la présente étude est d'évaluer l'efficacité de diverses méthodes de traitement des valeurs manquantes dans le contexte d'utilisation de données d'échelles de type Likert. L'efficacité est définie par la capacité d'une méthode de traitement de fournir des estimés de paramètres qui soient le plus près possible de la réelle valeur du paramètre de la population. Cette efficacité est évaluée en fonction de deux critères, soit un estimé de la fidélité de l'échelle de mesure, le coefficient alpha de Cronbach, ainsi qu'un estimé du trait latent mesuré par l'échelle, le score total classique. De plus, en manipulant certains facteurs, il est possible d'évaluer l'impact du nombre d'items, de la taille de la banque de données et de la proportion de valeurs manquantes.

Les résultats relatifs au coefficient alpha de Cronbach démontrent que la méthode d'imputation par la moyenne du sujet permet d'obtenir des estimés moins biaisés lorsque l'échelle de mesure compte quatre items, mais que pour l'échelle de neuf items, les méthodes de régression et d'imputation multiple semblent aussi efficaces que la méthode d'imputation par la moyenne du sujet. Il importe toutefois de souligner que la différence de biais entre la méthode d'imputation par la moyenne du sujet et les méthodes de régression et d'imputation multiple se situe à la deuxième décimale dans le cas de l'échelle de quatre items. Étant donné que la valeur du coefficient alpha de Cronbach se situe généralement entre 0 et 1, sur le plan pratique, une telle différence à la deuxième décimale suggère une efficacité pratiquement identique pour ces trois méthodes. La méthode d'élimination par liste, quant à elle, présente les valeurs moyennes de biais les plus élevées. Cette valeur peut même être supérieure à ,300 dans le cas de l'échelle de quatre items.

Les résultats de la présente étude confirment en partie ceux tirés des études antérieures, bien que certaines divergences existent entre les conclusions de ces études. En effet, Downey et King (1998) ainsi que McDonald, Thurston et Nelson (2000) concluent que les valeurs moyennes des biais étaient identiques peu importe les méthodes de traitement utilisées. D'un autre côté, Huisman (2000) et Enders (2003) suggèrent plutôt qu'il existe des différences entre les valeurs moyennes de biais des méthodes de traitement des valeurs manquantes. Les résultats obtenus par ces deux chercheurs indiquent que les méthodes d'imputation par la moyenne de l'item et d'élimination par liste présentent des valeurs de biais de l'estimé du coefficient alpha de Cronbach plus grandes que pour les autres méthodes évaluées par chacune des études : une conclusion qui corrobore en tout point celle obtenue dans la présente étude.

Autre point de concordance avec les études antérieures, la méthode de remplacement par la moyenne du sujet s'avère être une méthode efficace en ce qui concerne le coefficient alpha de Cronbach. Enders (2003) souligne que la valeur du coefficient alpha de Cronbach est moins biaisée lorsque la méthode d'imputation par la moyenne du sujet est utilisée, et qu'elle s'avère même plus efficace que la méthode EM (*Expectation-Maximization*) lorsque les données sont manquantes selon un mécanisme non aléatoire. Huisman (2000), quant à lui, révèle que la méthode d'imputation par la moyenne du sujet, bien que n'étant pas la méthode la moins biaisée, demeure une méthode efficace pour ce qui est de la réduction des biais causés par la présence de valeurs manquantes.

La bonne performance de la méthode d'imputation par la moyenne du sujet pour l'estimation du coefficient alpha de Cronbach soulève quelques interrogations. Enders (2003) supposait que l'observation de l'efficacité de cette méthode était due au fait que dans sa procédure de simulation des données, les items présentaient des moyennes identiques. Dans la situation où les items présentant des valeurs manquantes possèdent des moyennes significativement plus faibles que les autres items, il est attendu que la méthode d'imputation par la moyenne du sujet présentera des résultats beaucoup moins positifs que ceux obtenus dans son étude (Enders, 2003).

Selon Downey et King (2001), en utilisant la moyenne des autres items pour imputer une valeur manquante pour un item, cela aurait alors comme effet d'augmenter la covariance entre les items complets et l'item dont la valeur a été imputée. En conséquence, la méthode d'imputation par la moyenne du sujet procure une matrice de corrélations inter-items qui surestiment les valeurs des corrélations entre les items dont des données sont manquantes et les items dont les réponses sont complètes, procurant du même coup une valeur du coefficient alpha plus élevée que pour les autres méthodes de traitement.

Sur le plan du score à l'échelle, les résultats sont quelque peu divergents de ceux obtenus pour l'indice de fidélité. En effet, c'est la méthode d'élimination par liste qui permet d'obtenir les corrélations moyennes les plus élevées entre les scores après traitement des valeurs manquantes et ceux obtenus à partir des matrices de données complètes. La méthode d'imputation par la moyenne du sujet, quant à elle, présente des valeurs de corrélations moyennes plus petites. Cette valeur est même de ,878 pour l'échelle de quatre items lorsque la proportion de valeurs manquantes est de 25%. Il est à remarquer que les méthodes de régression et d'imputation multiple obtiennent des valeurs de corrélations moyennes légèrement inférieures à celles de la méthode d'élimination par liste.

La plupart des études qui ont abordé la notion de score à l'échelle dans leur évaluation suggèrent que les biais sur le score à l'échelle sont très peu importants, et ce, peu importe la méthode de traitement utilisée. Ce phénomène est observé autant pour les données basées sur un mécanisme de données effacées aléatoirement (Downey & King, 1998, McDonald, Thurston & Nelson, 2000, ou Roth, Switzer & Switzer, 1999) ou manquantes selon un mécanisme non aléatoire (Raaijmakers, 1999). De plus, tel que Downey et King (1998) l'ont observé, les corrélations entre le score obtenu à partir de l'application d'une méthode de traitement et le score à partir des données complètes sont plus petites pour la méthode d'imputation par la moyenne du sujet que pour les autres méthodes.

Concernant la question des scores à l'échelle, le mécanisme retenu pour créer les valeurs manquantes peut expliquer les petites différences observées entre les méthodes quant au recouvrement du score de l'échelle. Le mécanisme utilisé ciblait les individus possédant le score le plus faible à l'échelle sur les items présentant la moyenne la plus petite. Ainsi, la méthode d'imputation par la moyenne du sujet, qui utilise les résultats des autres items pour imputer une valeur à une donnée manquante, peut s'avérer efficace dans les conditions de l'étude. Il est donc fort probable que les individus ayant des valeurs manquantes présentaient en fait un score faible pour chaque item de l'échelle. Ainsi, en utilisant la moyenne des items répondus pour imputer une valeur aux items non répondus, la probabilité est tout de même assez élevée d'imputer la bonne valeur ou une valeur assez proche aux items non répondus.

Les résultats de l'étude suggèrent donc que les méthodes de traitement des valeurs manquantes les plus efficaces soient les méthodes d'imputation multiple et de régression. Malgré que ces méthodes présentent un biais légèrement supérieur à celui de la méthode d'imputation par la moyenne du sujet sur le plan de l'estimé du coefficient alpha de Cronbach, et ce, uniquement pour l'échelle de quatre items, cette différence est compensée par une meilleure efficacité sur le plan de l'estimation du score à l'échelle. La méthode d'imputation par la moyenne du sujet serait moins efficace que les méthodes d'imputation multiple et de régression, étant donné que les corrélations moyennes observées entre les scores obtenus pour la méthode d'imputation par la moyenne du sujet et ceux obtenus à partir des données complètes sont plus petites que celles des deux autres méthodes.

En ce qui a trait au deuxième objectif de l'étude, notons premièrement qu'il semble risqué de comparer les résultats obtenus selon le nombre d'items de l'échelle de mesure. Au départ, ces deux échelles ne présentaient pas des qualités psychométriques semblables. Par exemple, quant au coefficient alpha de Cronbach de la population, l'échelle de quatre items présente un coefficient de ,517 tandis que ce même coefficient atteint une valeur de ,873 pour l'autre échelle. La volonté de baser la présente étude sur des données réelles de recherche rendait pratiquement impossible la mission d'établir deux échelles de mesure ne présentant pas le même nombre d'items, mais présentant des coefficients alpha de Cronbach semblables.

Cependant, il est attendu que des méthodes d'imputation basées sur l'estimation de paramètres de modèles, telles que les méthodes d'imputation multiple ou de régression, soient plus efficaces lorsque le nombre d'items d'une échelle de mesure augmente. En effet, l'estimation des paramètres d'un modèle, comme un modèle de régression par exemple, sera d'autant plus précise si un plus grand nombre de données sont disponibles (Schaffer & Graham, 2002). L'effet de ce facteur mérite ainsi une attention plus approfondie.

Les résultats des ANOVA indiquent que les deux autres facteurs qui ont été considérés, soit la proportion de valeurs manquantes et la taille de l'échantillon, semblent avoir un impact peu important sur les valeurs moyennes des biais. Enders (2003) et Huisman (2000) arrivent à une conclusion identique sur l'impact de la taille de l'échantillon. Toutefois, Huisman (2000) suggère que la proportion de valeurs manquantes ait un impact sur l'efficacité des méthodes de traitement, ce qui est contraire aux résultats obtenus dans la présente étude. Toutefois, la proportion de valeurs manquantes semble avoir un effet plus important sur le score à l'échelle. En effet, les corrélations entre le score obtenu à partir d'une méthode de traitement et le score obtenu à partir des données complètes sont légèrement plus petites lorsque la proportion de valeurs manquantes est de 25%.

En réponse à la dernière question de recherche, bien que le nombre d'items semble avoir un impact sur l'efficacité des méthodes de traitement, il est impossible de départager l'influence du nombre d'items de celle des caractéristiques de ces items. Enfin, la proportion de valeurs manquantes semble avoir un certain impact sur l'efficacité des méthodes de traitement quant aux scores à l'échelle.

Conclusion

Les résultats de la présente étude sont quelque peu encourageants par rapport aux pratiques actuelles concernant le traitement des valeurs manquantes. Bien que la méthode qui est probablement la plus populaire, soit la méthode d'élimination par liste, présente des biais plus importants sur le plan de l'estimation du coefficient alpha de Cronbach, ce problème serait compensé par le fait que les biais sont les moins importants quant au score à l'échelle. Toutefois, les méthodes plus «modernes» telles que l'imputation multiple ou l'imputation par régression, présentent des résultats plus satisfaisants.

Cette étude met en relief l'importance de demeurer prudent lors de la conduite d'une étude d'évaluation des qualités psychométriques d'échelles de mesure. C'est par rapport à ce type d'étude que la présence de valeurs manquantes devrait intéresser les constructeurs d'échelles de mesure. Dans la situation où un nombre important de valeurs manquantes sont observées pour un item ou qu'un grand nombre de sujets n'ont pas répondu à certaines questions, il importe d'essayer de comprendre les raisons pouvant expliquer ce phénomène. Certaines décisions devront être prises pour les items dont la proportion de non-réponses est trop élevée.

Soulignons enfin certaines limites relatives à cette étude. Le fait d'utiliser un mécanisme simple d'effacement des données semble constituer une des principales limites à la généralisation des résultats de l'étude. Tel que l'ont souligné Little et Rubin (2002), le mécanisme causant la présence de valeurs manquantes est le facteur qui aurait l'impact le plus important sur l'efficacité des méthodes de traitement. Il faut donc faire preuve de prudence en ce qui concerne la généralisation des résultats de la présente étude. La plus grande efficacité dont fait preuve l'imputation multiple pourrait être attribuable à une simulation de mécanisme d'effacement des valeurs manquantes trop simple. À ce sujet, Yuan et Bentler (2000) exprimaient cette préoccupation en suggérant qu'il soit plus probable qu'une certaine proportion des données manquantes dans une matrice de données soit des données manquantes selon un mécanisme complètement aléatoire, d'autres manquantes selon un mécanisme aléatoire et enfin une dernière proportion serait des données manquantes selon un mécanisme non aléatoire. Il ne serait pas toutefois bien réaliste de tenter de déterminer correctement ces bonnes proportions pour les simuler.

Une deuxième limite est relative au nombre de méthodes sur lesquelles l'étude a porté. Comme il existe un grand nombre de méthodes de traitement des valeurs manquantes, il est impensable d'évaluer l'efficacité de toutes ces méthodes dans une seule étude. Un certain choix de méthodes a donc dû être effectué pour la réalisation de l'étude. Ainsi, bien que l'imputation multiple semble être une méthode efficace relativement aux autres méthodes évaluées, d'autres méthodes de traitement auraient pu se montrer supérieures.

Malgré que les connaissances en ce domaine ne soient pas encore très développées, certaines recommandations pour la pratique en recherche peuvent tout de même être avancées. À l'instar d'Allison (2001), il importe de rappeler qu'il faut tout faire pour réduire au minimum la présence des valeurs manquantes. Comme le dit si bien cet auteur, la meilleure façon de traiter les valeurs manquantes, c'est encore de viser à ne pas en avoir du tout. Il est aussi fortement suggéré d'effectuer une première analyse exploratoire des données dans le but de prendre conscience de l'ampleur du problème dans la banque de données. Cette analyse des données permettrait d'évaluer les biais probables qui risquent d'être encourus dans les analyses subséquentes.

En conclusion, puisqu'il est effectivement impossible d'éliminer tous les biais causés par les valeurs manquantes (Allison, 2001), il importe de toujours demeurer conscient que les banques de données de recherche sont incomplètes et qu'une méthode de traitement des valeurs manquantes a bien souvent dû être appliquée avant d'effectuer les analyses statistiques, avec comme conséquence, qu'on le veuille ou non, que les résultats de toute recherche peuvent être en partie attribuables à la méthode de traitement choisie pour traiter les données manquantes.

RÉFÉRENCES

- Allison, P.D. (2001). *Missing data*. Sage University papers series on quantitative applications in the social sciences, 07-136. Thousand Oaks, CA: Sage.
- Downey, R.G., & King, C.V. (1998). Missing data in Likert rating: A comparison of replacement methods. *Journal of General Psychology, 125*, 175-191.
- Enders, C.K. (2003). Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. *Psychological Methods, 8*, 322-337.
- Enders, C.K. (2004). The impact of missing data on sample reliability estimates: Implications for reliability reporting practices. *Educational and Psychological Measurement, 64*, 419-436.
- Fichman, M., & Cummings, J.N. (2003). Multiple imputation for missing data: Making the most of what you know. *Organizational Research Methods, 6*, 282-308.

- Green, S.B., & Salkind, N.J. (2003). *Using SPSS for Windows and Macintosh: Analyzing and understanding data*. Prentice Hall: New Jersey.
- Huisman, M. (2000). Imputation of missing item responses: Some simple techniques. *Quality and Quantity*, 34, 331-351.
- Little, R.J.A., & Rubin, D.B. (2002). *Statistical analysis with missing data*, 2nd edition. New Jersey: John Wiley & Sons Inc.
- McDonald, R.A., Thurston, P.W., & Nelson, M.R. (2000). A Monte Carlo study of missing item methods. *Organizationnal Research Methods*, 3, 70-91.
- Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London, Series A*, 236, 333-380.
- Neyman, J., & Pearson, E.S. (1933). On the problem of most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London, Series A*, 231, 289-337.
- Raaijmakers, Q.A.W. (1999). Effectiveness of different missing data treatments in surveys with Likert-type data: Introducing the relative mean substitution approach. *Educational and Psychological Measurement*, 59, 725-748.
- Roth, P.L., Switzer, F.S., & Switzer, D.M. (1999). Missing data in multiple item scales: A Monte Carlo analysis of missing data techniques. *Organizational Research Methods*, 2, 211-232.
- Rubin, D.B. (1976). Inference and Missing Data, *Biometrika*, 29, 159-183.
- Schaffer, J.L. (1998). *Analysis of incomplete multivariate data*. New York: Chapman & Hall.
- Schaffer, J.L., & Graham, J.W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147-177.
- Sinharay, S., Stern, H.S., & Russel, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods*, 6, 317-329.
- van der Kamp, L.J.T., & Bijleveld, C.C.J.H. (1998). Methodological issues in longitudinal research. In C.C.J.H. Bijleveld & L.J.T. van der Kamp, *Longitudinal data analysis* (pp. 1-45). London: Sage Publications Ltd.
- Yuan, K.-H., & Bentler, P.M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. In M. Becker & M. Sobel (éds), *Sociological methodology 2000*. Malden, MA: Blackwell.