Document généré le 29 mars 2024 08:07

Mesure et évaluation en éducation

Validation de la version française de la *Behavior and Instructional Management Scale*

Philippe Wanlin et Malika Revilloud

Volume 38, numéro 2, 2015

Réception : 12 juillet 2013
Version finale : 28 janvier 2015
Acceptation : 16 février 2015

URI : https://id.erudit.org/iderudit/1036763ar
DOI : https://doi.org/10.7202/1036763ar

Aller au sommaire du numéro

Éditeur(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (imprimé)
2368-2000 (numérique)

Découvrir la revue

Citer cet article

Wanlin, P. & Revilloud, M. (2015). Validation de la version française de la *Behavior and Instructional Management Scale*. *Mesure et évaluation en éducation*, *38*(2), 61–85. https://doi.org/10.7202/1036763ar

Résumé de l'article

La Behavior and Instructional Management Scale (BIMS) développée par Martin et Sass (2010) permet de mesurer les pratiques et les croyances des enseignants en gestion de la classe. Elle est composée de deux dimensions théoriques : la gestion de l'enseignement et la gestion du comportement des élèves. Martin et Sass (2010) ont validé la version anglaise de cet instrument de mesure. L'objectif de cet article est de présenter la chronologie de son évolution, puis de valider sa version française auprès de 338 enseignants qui se forment à l'enseignement au primaire et au secondaire. Les résultats, qui sont discutés en termes de perspectives de recherche, montrent que la traduction française de la BIMS obtient des indicateurs de qualités psychométriques proches de ceux rapportés par Martin et Sass (2010).

Tous droits réservés © ADMEE-Canada - Université Laval, 2015

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne.

https://apropos.erudit.org/fr/usagers/politique-dutilisation/

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

https://www.erudit.org/fr/

Validation de la version française de la *Behavior and Instructional Management Scale*

Philippe Wanlin
Université de Genève (IUFE)

Malika Revilloud
Université de Genève (IUFE-FPSE)

Mots clés : gestion de la discipline, gestion du comportement des élèves, gestion de l'enseignement, instrument de mesure

La Behavior and Instructional Management Scale (BIMS) développée par Martin et Sass (2010) permet de mesurer les pratiques et les croyances des enseignants en gestion de la classe. Elle est composée de deux dimensions théoriques : la gestion de l'enseignement et la gestion du comportement des élèves. Martin et Sass (2010) ont validé la version anglaise de cet instrument de mesure. L'objectif de cet article est de présenter la chronologie de son évolution, puis de valider sa version française auprès de 338 enseignants qui se forment à l'enseignement au primaire et au secondaire. Les résultats, qui sont discutés en termes de perspectives de recherche, montrent que la traduction française de la BIMS obtient des indicateurs de qualités psychométriques proches de ceux rapportés par Martin et Sass (2010).

Key words: behavior management, instructional management, measurement scale

Martin and Sass (2010) have construct-validated the Behavior and Instructional Management Scale (BIMS), an instrument that facilitates assessment of teachers' beliefs and practices in classroom management. This scale is subdivided into two independent theoretical constructs: instructional management and behavior management. Our aim is to report on the evolution of the scale and to validate its French translated version. In this perspective, we collected data from a sample containing 338 pre-service elementary and secondary school graduating teachers. Results show that the French translated version of the BIMS obtains psychometric properties that are close to those observed by Martin and Sass (2010). These results are discussed in terms of research perspectives.

Palavres-chave: gestão da disciplina, gestão das abordagens pedagógicas, instrumento de medida

A Behavior and Instructional Management Scale (BIMS) desenvolvida por Martin e Sass (2010) permite medir as práticas e as crenças dos professores na gestão da sala de aula. Trata-se de uma escala composta por duas dimensões teóricas: a gestão do ensino e a gestão dos comportamento dos alunos. Martin e Sass (2010) validaram a versão anglófona deste instrumento de medida. O objetivo deste artigo é validar a sua versão francesa, apresentando também a cronologia da sua evolução. Neste sentido, recolheram-se dados de uma amostra de 338 professores em formação inicial para o ensino básico e secundário. Os resultados, que são discutidos em termos de perspectivas de investigação, mostram que a tradução francesa do BIMS obtém indicadores com qualidades psicométricas próximas das observadas por Martin e Sass (2010).

Note des auteurs: La correspondance liée à cet article peut être adressée à Philippe Wanlin, téléphone: +41 22 379 04 27, mais de préférence à l'adresse courriel suivante: [philippe.wanlin@unige.ch].

Les auteurs remercient les responsables de la Section des sciences de l'éducation de la Faculté de psychologie et des sciences de l'éducation de l'Université de Genève pour leur soutien financier.

Introduction

La *Behavior and Instructional Management Scale* (BIMS) de Martin et Sass (2010) est un instrument de mesure des pratiques déclarées des enseignants concernant leur gestion de l'enseignement (*instructional management*) et du comportement des élèves en classe (*behavior management*). Cet outil est le produit d'un travail s'inscrivant dans la continuité des outils de mesure créés et testés par Martin et ses collaborateurs depuis le début des années 1990.

Dans cet article sera retracée, dans les grandes lignes, l'évolution des échelles développées en anglais par Martin et ses collaborateurs, à commencer par l'instrument qui a amorcé leurs travaux pour terminer par l'outil le plus récent. Cette partie théorique sera suivie d'une partie empirique qui décrit une étude de validation de la version française de la BIMS auprès de 338 enseignants en formation. À la suite des analyses factorielles exploratoire et confirmatoire ainsi que des analyses de consistance interne menées pour asseoir les qualités psychométriques de cet instrument, les résultats seront discutés en termes de perspectives de recherche et de potentialités de la BIMS pour évaluer l'évolution des compétences des enseignants en matière de gestion des stratégies d'enseignement déployées pour favoriser l'apprentissage, et de gestion de la discipline et du comportement des élèves en classe.

Chronique du développement de la BIMS

Les instruments qui ont précédé la BIMS

Comme le montre le tableau 1, tout a commencé par l'*Inventory of Classroom Management Style* (ICMS), qui mesure les perceptions des enseignants sur la gestion de la classe (Martin & Baldwin, 1993a). Ce premier outil est une adaptation du *Beliefs on Discipline Inventory* (BDI) (Glickman & Tamashiro, 1980 ; Wolfgang & Glickman, 1986). Cet outil propose des affirmations et deux possibilités de réponse pour chacune d'elles qui renvoient à une des trois tendances managériales postulées par Glickman et ses collaborateurs (relations humaines et écoute ; clarification des règles avec aide à l'autogestion ; et approche behavioriste du

management combinant règles, récompenses et sanctions). L'enseignant peut calculer sa tendance managériale en comptant le nombre d'affirmations correspondant à chaque tendance, celle contenant la somme la plus élevée correspondant à ses conduites et à ses croyances dominantes.

À la base, le BDI traite des croyances des enseignants en matière de discipline et mesure donc trois styles d'interaction au sein de la classe, sur un continuum dont le premier extrême est le style interventionniste. Ce style renvoie à ce que fait l'environnement extérieur pour permettre à un individu de se développer d'une manière particulière. À l'autre extrême, le style non interventionniste présuppose que l'élève a une force intérieure qui doit trouver son expression dans le monde réel. Au centre du continuum se situe le style interactionniste, qui fait référence aussi bien à ce que l'individu doit faire pour modifier l'environnement extérieur qu'à ce que l'environnement fait pour façonner l'individu (Wolfgang & Glickman, 1986). L'hypothèse est qu'un enseignant ayant un style interventionniste est plus directif dans sa gestion de classe, et inversement pour le style non interventionniste. Martin et Baldwin (1993a) conservent cette base, mais adaptent l'outil en ajoutant des oppositions d'affirmations.

Inspiré de ce continuum, l'ICMS mesure les trois dimensions qui constituent la gestion de la classe selon Martin et Baldwin (1993a) : ce qui se rapporte à l'élève ou, plus précisément, ce que l'enseignant fait pour permettre à l'élève de se développer en tant que personne ; ce qui se rapporte à l'enseignement, c'est-à-dire ce que l'enseignant fait pour permettre à l'élève d'apprendre ; et ce qui se rapporte à la discipline, à savoir les comportements de l'enseignant pour établir des normes et ceux pour les faire respecter. Cet instrument est composé d'un total de 48 items, dont 12 pour la dimension « personne », 24 pour la dimension « enseignement » et 12 pour la dimension « discipline ». Les items sont composés de paires d'affirmations opposées et les enseignants sont invités à choisir l'affirmation qui leur correspond le plus. Chaque affirmation est associée à un score et renvoie à un style managérial : les affirmations interventionnistes donnent 2 points, tandis que les affirmations non interventionnistes donnent 1 point. L'addition de ces scores d'affirmations donne un score total à chaque individu sur le continuum non interventionniste (24 points)/interactionniste (36 points)/interventionniste (48 points). Martin et Baldwin (1993a) ont demandé à 158 enseignants, dont 96 en formation, de remplir leur questionnaire. Concernant les indicateurs de qualité, seul un alpha

Tableau 1

Tableau synoptique des études portant sur le développement de la BIMS

Étude	Instrument	Échantillon	Nombre d'items total	Nombre d'items disponibles	Échelles	α	Score de management
Martin & Baldwin (1993a)	ICMS	96* 62	48	12/48	Choix forcé entre deux propositions opposées — dimension personne (6 it. = 12 op.) — dimension enseignement (12 it. = 24 op.) — dimension discipline (6 it. = 12 op.)	0,6985	Score entre 24 (non interventionniste) et 48 (interventionniste); 36 idéologie interactionniste
Martin & Baldwin (1993b)	ICMS	238*	48	12/48	Likert à 4 modalités — dimension personne (12 it.) — dimension enseignement (24 it.) — dimension discipline (12 it.)	0,6985	Score entre 48 (non interventionniste) et 192 (interventionniste); 120 idéologie interactionniste
Martin & Baldwin (1994)	ICMS	131* 107	41	n. i.	Likert à 4 modalités	0,64	Score entre 41 (non interventionniste) et 164 (interventionniste); 102,5 idéologie interactionniste
Martin & Yin (1997)	ABCC	282	29	n. i.	Likert à 4 modalités — dimension personne (9 it.) — dimension enseignement (14 it.) — dimension comportement (6 it.)	 0,7345 0,8220 0,6523	Score unique
Martin, Yin & Baldwin (1998a)	ABCC	282	48	48	Likert à 4 modalités — dimension personne (8/9 it.) — dimension enseignement (14/24 it.) — dimension comportement (4/15 it.)	 0,69 0,82 0,69	Un score par sous-échelle (moyenne des scores d'items)
Martin, Yin & Baldwin (1998b)	ABCC	281	26	n. i.	Likert à 4 modalités — dimension personne (9 it.) — dimension enseignement (14 it.) — dimension comportement (6 it.)	 0,69 – 0,82	Score unique sur un continuum non interventionniste/interventionniste (avec au centre le style interactionniste)
Martin, Yin & Mayall (2007)	ABCC-R	489	20	50/20	Likert à 4 modalités — dimension personne et comportement (10 it.) — dimension enseignement (10 it.)	 0,77 0,78	Score unique sur un continuum non interventionniste/interventionniste (avec au centre le style interactionniste)

n. i. = non indiqué dans le texte de référence * étudiants en éducation it. = items op. = oppositions, affirmations contradictoires

ICMS = *Inventory of Classroom Management Style*
ABCC = *Attitudes and Beliefs on Classroom Control Inventory*
ABCC-R = *Revised Attitudes and Beliefs on Classroom Control Inventory*

de Cronbach d'une valeur proche de 0,7 pour l'échelle totale est proposé, bien que trois dimensions soient postulées. Les auteurs procèdent ensuite à des analyses de la variance (ANOVA) et utilisent des coefficients de corrélation pour valider leur outil.

La même année, l'instrument subit une autre transformation : les oppositions sont remplacées par un choix entre quatre degrés d'accord : me décrit parfaitement, me décrit généralement, me décrit à peu près et ne me décrit pas du tout (Martin & Baldwin, 1993b). Les auteurs ont demandé à 238 enseignants en formation de remplir ce nouveau questionnaire, qui obtient un alpha de Cronbach d'environ 0,7. Une analyse en composantes principales avec rotation varimax aboutit à environ 30 % de variance expliquée pour une solution en trois composantes. La première composante agrège 13 items, la deuxième, 17 et la troisième, 10. Pour Martin et Baldwin (1993b), leur instrument est imparfait, car des items saturent dans des composantes dans lesquelles ils ne devraient pas d'un point de vue théorique ; ce qui signifie, à leur avis, que la gestion de l'enseignement est constituée de facettes multiples inextricablement liées. Cependant, l'utilisation d'une analyse en composantes principales serait non adaptée pour la validation d'échelles multidimensionnelles (Furr, 2011) et l'absence des indicateurs d'ajustement générés par une analyse factorielle confirmatoire nécessaires à la vérification de la qualité effective de l'outil est à déplorer.

Martin et Baldwin (1994) proposent une analyse des différences novices-experts en matière de gestion de l'enseignement et de la discipline à l'aide de l'ICMS. Pour cette étude, ils conservent 41 items de la version antérieure de l'ICMS, celle où les participants doivent marquer leur degré d'accord avec chacun d'eux sur une échelle de Likert à quatre modalités. Les individus obtiennent également un score global allant de 41 (non interventionniste) à 164 (interventionniste). Ils obtiennent, pour l'ensemble des items, une consistance interne de 0,64 et procèdent surtout à des analyses de variance pour comparer les scores de non-interventionnisme ou d'interventionnisme des novices et des experts. Ils montrent globalement que ces premiers sont plus interventionnistes que ces derniers. Il est à nouveau regrettable que les auteurs n'aient pas validé leur outil comme il se doit, à savoir en proposant une analyse factorielle confirmatoire.

L'ICMS a ensuite été rebaptisé *Attitudes and Beliefs on Classroom Control Inventory* (ABCC), dans lequel seuls 29 items sont conservés (Martin & Yin, 1997). Ces auteurs proposent alors trois indicateurs de consistance interne à l'aide de leurs données reposant sur les réponses de 282 enseignants titulaires, dont 172 exercent dans l'enseignement secondaire. La valeur des alphas est de 0,82 pour l'échelle gestion de l'enseignement, de 0,73 pour la gestion des personnes et de 0,65 pour la gestion de la discipline. Hélas, les auteurs conservent la composition d'un score unique opposant non-interventionnisme et interventionnisme. Leur validation repose encore sur des ANOVA. Martin, Yin et Baldwin (1998a) réduisent l'instrument à 26 items. Le questionnaire est soumis à 281 enseignants titulaires, qui obtiennent des alphas de 0,82 à 0,69. Ces auteurs procèdent exactement aux mêmes comparaisons ANOVA sur un score unidimensionnel. Martin, Yin et Baldwin (1998b) reprennent ces données pour obtenir des résultats similaires en adoptant des standards d'analyse légèrement différents.

L'ABCC a ensuite été utilisé par d'autres chercheurs (par ex., Ritter & Hancock, 2007 ; Sokal, Smith, & Mowat, 2003 ; Yilmaz & Çavaş, 2008). Aucune des recherches relevées n'a validé l'outil à l'aide d'une analyse factorielle confirmatoire. Sokal et ses collaborateurs (2003) ont sollicité 82 enseignants en formation pour répondre aux 26 items reposant sur les trois dimensions théoriques et ont calculé un score unidimensionnel pour comparer, par ANOVA, les participants en fonction de leur avancement dans leur formation. Ritter et Hancock (2007) présentent des données rassemblées auprès de 158 enseignants (dont 60 suivent une formation initiale alternative). Ils rapportent des alphas entre 0,69 et 0,82 pour les trois dimensions théoriques et écrivent avoir procédé à une analyse factorielle exploratoire, mais ne la présentent pas dans leur article. Ils calculent aussi un score unidimensionnel et procèdent à des ANOVA. Yilmaz et Çavaş (2008) ont interrogé 185 enseignants en formation. Ils conservent 23 des 26 items pour une solution en deux dimensions par le biais d'une analyse en composantes principales. Le premier facteur est composé de 12 items et explique 13,87 % de variance, tandis que le second facteur rassemble 11 items pour expliquer 19,78 % de la variance. Leurs alphas sont bons pour les deux facteurs qui renvoient à la gestion de la personne (0,78) et à la gestion de l'enseignement (0,74). L'ensemble de l'échelle obtient un alpha de 0,75. Enfin, ils proposent des comparaisons par ANOVA sur un score unidimensionnel.

Tenant compte des résultats d'autres études (Henson, 2003 ; Savran & Cakiroglu, 2003), Martin, Yin et Mayall (2007) proposent une révision de l'ABCC : l'ABCC-R. La version révisée conserve 20 items et réduit la mesure à deux dimensions, soit la gestion de l'enseignement et la gestion de la personne. (La gestion du comportement est mixée à cette dernière à cause de sa faiblesse par rapport aux deux autres dimensions.) Les alphas de Cronbach pour ces deux dimensions sont proches de 0,80 et les 489 enseignants titulaires interrogés obtiennent à nouveau un score unique sur un continuum non interventionniste/interventionniste. Hélas, les auteurs ne soumettent pas leur instrument à une validation par analyse factorielle confirmatoire.

En somme, l'instrument de Martin et ses collaborateurs a subi des transformations importantes au fur et à mesure des années. D'abord, la banque d'items s'est étoffée. D'une organisation en affirmations opposées, l'instrument est passé à une échelle de Likert à quatre modalités, puis la mesure a été réduite de trois à deux dimensions. D'autres éléments sont remarquablement stables : un score unique pour chaque participant sur un continuum non interventionniste/interventionniste qui se traduit, en règle générale, par le calcul d'un alpha de Cronbach pour l'ensemble de l'échelle. Néanmoins, la traduction de plusieurs dimensions en un score unique pose problème, puisque cela ne concorde pas avec le postulat théorique sous-jacent de l'existence de deux à trois dimensions de management de la classe. Ce postulat devrait donc être vérifié par des analyses psychométriques adéquates, d'autant plus que les analyses factorielles exploratoires suggèrent au moins deux facteurs et que les chercheurs mentionnent, dans certains cas, un coefficient de consistance interne pour chacun d'eux. Un autre problème est qu'aucune des références ne mentionne d'indicateurs d'ajustement des données au modèle attestant la validité des outils, et il en va de même pour les recherches qui les ont utilisés (par ex., Ritter & Hancock, 2007 ; Sokal, Smith, & Mowat, 2003 ; Yilmaz & Çavaş, 2008).

La validation de la BIMS

Inspirée de ces différents instruments de mesure, la BIMS (Martin & Sass, 2010) est composée de deux dimensions théoriques. La gestion du comportement (*behavior management – BM*) se rapproche de la discipline, et renvoie aux efforts de l'enseignant pour prévenir les mauvais comportements et pour y répondre à travers l'établissement de règles, d'un systè-

me de récompenses et de possibilités de participation des élèves (Martin & Sass, 2010). La gestion de l'enseignement (*instructional management – IM*) fait référence aux objectifs éducatifs et aux méthodes des enseignants. Elle inclut les routines quotidiennes et le suivi du travail en classe, mais également le type d'enseignement, c'est-à-dire l'instruction frontale directe et l'approche participative ou coopérative (Martin & Sass, 2010). L'outil initial est composé de 24 items, dont 12 par dimension théorique, créés d'après l'hypothèse que ces deux dimensions transcendent l'enseignement. Les participants se positionnent sur ces affirmations grâce à six degrés d'accord. Même s'ils ne les calculent pas, Martin et Sass (2010) expliquent que leur outil permet de générer deux scores : le premier concerne la gestion de la discipline, où un score faible correspond au non-interventionnisme et un score élevé à l'interventionnisme, tandis que le second renvoie à l'enseignement, où un score bas correspond aux approches constructivistes et un score élevé à l'enseignement transmissif. Martin et Sass (2010) suggèrent de générer ce score soit en prenant la moyenne des degrés d'accord des participants sur les items composant les deux dimensions théoriques, soit en utilisant les scores factoriels obtenus lors des analyses factorielles.

Ils soumettent ces 24 affirmations à un échantillon de 550 enseignants, qu'ils décomposent en deux ensembles aléatoirement constitués : 200 enseignants pour une première analyse exploratoire (étude 1), puis 350 pour une analyse confirmatoire (étude 2). Ils proposent enfin une troisième étude confirmatoire sur les 550 participants (étude 3), auxquels ils ont aussi soumis l'*Ohio State Teacher Efficacy Scale* (OSTES) (Tschannen-Moran & Woolfolk-Hoy, 2001).

Pour l'étude 1, Martin et Sass (2010) ont mené une analyse factorielle exploratoire (*exploratory factor analysis – EFA*) sur les données de 200 participants, choisis de manière aléatoire dans le but de raccourcir de moitié l'échelle de 24 items. Pour ce faire, ils utilisent le logiciel de statistiques M*plus* et ses fonctionnalités de gestion des données manquantes (FIML) et de traitement de données catégorielles (WLSMV avec rotation CF-Equamax). La version initiale de 24 items obtient d'intéressants indicateurs[1] de qualité d'ajustement pour une solution à deux facteurs ($\chi^2_{(229)}$ = 2749,352, $p < 0{,}001$; CFI = 0,934 ; TLI = 0,921 ; RMSEA = 0,098 ; SRMR = 0,072). Trois critères président leur procédure de raccourcissement de leur échelle initiale : 1) l'élimination des items ne saturant pas dans leur

facteur théorique mais dans l'autre facteur, 2) la conservation des items saturant à plus de 0,40 dans leur facteur théorique, et 3) l'élimination des items saturant à plus de 0,30 dans l'autre facteur. Cette procédure leur permet de conserver 12 items, soit six par dimension, qui composent la version courte de la BIMS. D'après ces auteurs, les valeurs de qualité d'ajustement témoignent de l'adéquation entre le modèle observé et le modèle théorique ($\chi^2_{(43)} = 92{,}153$, $p < 0{,}001$; CFI = 0,978 ; TLI = 0,966 ; RMSEA = 0,076 ; SRMR = 0,050). Le premier facteur, composé des items BM1, 2, 8, 9, 11 et 12, explique 31,8 % de la variance totale, tandis que le second facteur, composé des items IM2, 3, 5, 6, 9 et 12, en explique 22,84 %. La corrélation interfacteurs est relativement peu élevée, avec une valeur de 0,22.

L'analyse du tableau de saturation de Martin et Sass (2010) met en évidence qu'ils n'ont pas respecté à la lettre leur procédure d'écartement/conservation des items. Ainsi, pour le facteur BM, ils gardent l'item BM1, qui sature à 0,53, et rejettent les items BM7 et BM10, qui saturent respectivement à 0,58 et 0,56. Ceci est étrange, puisqu'une autre de leur règle consistait à éliminer les items par ordre de grandeur s'ils dépassaient 0,40 de saturation dans leur facteur théorique afin de ne conserver que six items par facteur. Quant au facteur IM, l'item IM7 a été ecarté malgré sa saturation de 0,60. Cette remarque aura une certaine importance plus loin dans le texte.

Pour l'étude 2, Martin et Sass (2010) procèdent à une analyse factorielle confirmatoire sur le reste de leur échantillon, à savoir 350 participants. Ils utilisent les mêmes fonctionnalités dans le logiciel M*plus*. Les valeurs d'ajustement des données au modèle théorique sont intéressantes ($\chi^2_{(28)} = 126{,}271$, $p < 0{,}001$; CFI = 0,932 ; TLI = 0,949 ; RMSEA = 0,10 ; WRMR = 1,142). Néanmoins, le logiciel M*plus* suggère de corréler les termes d'erreurs des items BM9 et IM9 qui, d'après la procédure DIFFTEST, permettent d'améliorer significativement la qualité d'ajustement du modèle ($\chi^2_{(28)} = 106{,}637$, $p < 0{,}001$; CFI = 0,945 ; TLI = 0,959 ; RMSEA = 0,090 ; WRMR = 1,040), mais qui, de l'aveu des auteurs, n'ont pas de pertinence théorique.

Pour l'étude 3, l'analyse factorielle confirmatoire appliquée sur les données des 550 participants conduit à un ajustement satisfaisant pour la version sans corrélation entre termes d'erreurs ($\chi^2_{(30)} = 186{,}402$, $p < 0{,}001$; CFI = 0,927 ; TLI = 0,947 ; RMSEA = 0,097 ; WRMR = 1,342), qui s'amé-

liore significativement en corrélant les termes d'erreurs de BM9 et IM9 ($\chi^2_{(30)}$ = 162,576, p < 0,001 ; CFI = 0,938 ; TLI = 0,953 ; RMSEA = 0,092 ; WRMR = 1,262). Les alphas de Cronbach témoignent d'une relativement bonne consistance interne des deux sous-échelles : 0,774 pour la sous-échelle BM et 0,770 pour la sous-échelle IM. L'article de Martin et Sass (2010) présente ensuite un modèle général combinant la BIMS et l'OSTES, mais qui ne sera pas traité ici, car il ne correspond pas à l'objectif de cette contribution de valider la version française de la BIMS.

Il faut saluer l'effort de Martin et Sass (2010), qui présentent un outil intéressant de mesure des pratiques déclarées des enseignants en matière de gestion de l'enseignement et du comportement des élèves en classe. Selon les standards en vigueur pour le jugement de la qualité d'ajustement des modèles, il faut cependant souligner que les auteurs sont généreux dans leur évaluation de leur instrument. En effet, avec des RMSEA oscillant autour de 0,09 et des CFI aux alentours de 0,93, les résultats sont relativement éloignés du non-dépassement de 0,05 pour la RMSEA et du dépassement de 0,95 pour le CFI.

Ce constat ainsi que l'invitation formulée par Martin et Sass (2010) à tester l'adéquation de leur modèle théorique dans d'autres contextes ont motivé l'étude rapportée dans cet article. L'objectif est de soumettre l'instrument de 24 items à un échantillon d'enseignants en formation afin de vérifier si nous aboutissons aux mêmes constats que Martin et Sass (2010) en adoptant exactement les mêmes procédures.

Méthode

Les participants

L'échantillon de cette étude est composé de 338 enseignants en formation, dont 139 se forment à l'enseignement au primaire et 197 à l'enseignement au secondaire (deux participants ne mentionnent pas l'information). Cet échantillon comprend 221 femmes, dont 110 suivent la formation à l'enseignement au primaire. Ces 338 enseignants en formation proviennent tous de l'Institut universitaire de formation des enseignants (IUFE) de l'Université de Genève. L'échantillon contient, pour l'enseignement primaire, les trois dernières années de la formation en quatre ans, excluant de ce fait la première année, qui est un tronc commun

également suivi par de futurs psychologues ou orthophonistes. Pour l'enseignement secondaire, l'échantillon est composé d'étudiants des deux années de formation.

La procédure et le traitement des données

Les enseignants en formation ont été sollicités lors des cours à l'IUFE pour remplir un questionnaire papier. Ce dernier est composé de la traduction en français des 24 items de Martin et Sass (2010), qui est présentée en annexe, mais également d'items d'autres instruments qui ne seront pas mentionnés ici (total de 96 items). Les participants étaient invités à indiquer leur degré d'accord à l'aide de six modalités, comme dans la version originale : *totalement d'accord* (TA), *d'accord* (A), *plutôt d'accord* (PA), *plutôt en désaccord* (PD), *en désaccord* (D) et *total désaccord* (TD).

Comme dans Martin et Sass (2010), l'analyse a utilisé, dans le logiciel M*plus* (Muthén & Muthén, 2002), un algorithme basé sur l'estimateur WLSMV (*weighted least squares mean- and variance-adjusted*) adapté à des données catégorielles. Une rotation CF-Equamax a été appliquée lors de l'analyse factorielle exploratoire. La gestion des données manquantes a été assurée, tant pour l'analyse factorielle exploratoire que confirmatoire, grâce à la procédure FIML (*full information maximum likelihood*), qui s'enclenche par défaut dans M*plus*. À noter que les données manquantes de cette étude sont peu nombreuses (<1 %).

Une analyse factorielle exploratoire a été menée sur la version longue de la BIMS, puis sur la version courte proposée par Martin et Sass (2010) afin de vérifier si on retrouve les mêmes valeurs que ces chercheurs. Ensuite, à partir de l'analyse factorielle exploratoire de la version longue, l'instrument a été raccourci selon la même procédure que Martin et Sass (2010). La version courte de Martin et Sass (2010) ainsi que celles retenues à la suite de l'application de leur procédure aux données de cette étude sont toutes soumises à une analyse factorielle confirmatoire.

Les indicateurs de qualité sont interprétés selon les standards définis par Brown (2006), qui s'inspire de Hu et Bentler (1999) et de Browne et Cudek (1993), et par ceux retenus par Muthén et Muthén (2002). Les valeurs critiques sont les suivantes :

– RMSEA : <0,10 = ajustement médiocre, <0,08 = ajustement adéquat, <0,06 = ajustement raisonnablement bon, <0,05 = bon ajustement ;

- SRMR : <0,08 = ajustement relativement bon (M*plus* génère cet indicateur uniquement pour l'analyse factorielle exploratoire ; idéalement, il devrait adopter une valeur ≤0,07.) ;
- CFI et TLI : <0,90 = modèle suspect à rejeter, 0,90< TLI/CFI <0,95 = ajustement acceptable, >0,95 = bon ajustement ;
- WRMR : ≤1,00 = modèle au bon ajustement. (M*plus* génère cet indicateur uniquement pour l'analyse factorielle confirmatoire opérée avec l'estimateur WLSMV à la place du SRMR.)

À noter que Brown (2006) souligne que l'analyse du χ^2, qui doit être non significatif (<0,05) pour que le modèle soit adéquat, ne doit pas empêcher de conserver un modèle donné. En effet, il suggère que le χ^2 est un indicateur historique, mais qu'il a le défaut de dépendre du nombre d'individus contenus dans l'échantillon. En cas d'échantillon de grande taille, le χ^2 tendrait inévitablement à être significatif (voir aussi Muthén et Muthén, 2002). Concernant le WRMR, les auteurs ne parviennent pas à trouver de consensus sur le fait qu'il doive se situer en dessous ou au-dessus de 1 (Wang & Wang, 2012 ; Muthén & Muthén, 2002). Par conséquent, Muthén et Muthén suggèrent, sur le site Internet[2] du logiciel M*plus*, de ne pas accorder trop de poids à cet indicateur expérimental, car il manque d'études pour décider de son critère d'évaluation.

La consistance interne des sous-échelles a été calculée de deux manières : le classique alpha de Cronbach et, étant donné les mises en garde de Laveault (2012) ainsi que l'invitation de Furr (2011) et de Brown (2006) à abandonner cet indicateur, un coefficient ajusté. Le calcul de cette valeur de fidélité ajustée est basé sur la formule[3] suivante : le carré de la somme des saturations factorielles (*factor loadings*) divisé par la somme de trois éléments, qui sont : 1) le carré de la somme des saturations factorielles, 2) la somme des erreurs de variance, et 3) le double de la somme des covariances entre les termes d'erreurs de certains items. (Ce dernier élément prend la valeur de 0 quand il n'y a pas de corrélation entre des termes d'erreurs.)

Résultats

L'analyse factorielle exploratoire appliquée sur les données qui concernent 338 enseignants en formation mène, pour la version longue, à un ajustement intéressant entre données et modèle théorique ($\chi^2_{(229)} = 628{,}934$, $p < 0{,}001$; CFI = 0,906 ; TLI = 0,887 ; RMSEA = 0,072 ; SRMR = 0,061).

Lorsque la version courte suggérée par Martin et Sass (2010) est retenue, ces valeurs attestent d'une raisonnablement bonne adéquation des données au modèle théorique ($\chi^2_{(43)}$ = 123,561, p < 0,001 ; CFI = 0,956 ; TLI = 0,933 ; RMSEA = 0,074 ; SRMR = 0,044). Le tableau 2, qui présente les saturations factorielles, indique que l'application de la version courte de Martin et Sass (2010) a un certain sens, puisque l'utilisation de la même procédure que ces auteurs aboutit à l'élimination de presque les mêmes items pour constituer la version courte (BIMS). Si on s'en tient strictement à leur procédure, il y a deux choix. Premièrement, si on veut garder six items par dimension, il faut remplacer IM3 par IM7 et BM8 par BM7 (BIMS-12). Rappelons que ces deux items obtenaient de très bonnes saturations dans l'étude de Martin et Sass (2010), mais que les auteurs les avaient écartés pour une raison indéterminée. Une autre solution consiste à éliminer les items BM8 et IM3 si on veut garder une certaine proximité à leur solution, bref de conserver une échelle ayant cinq items par dimension (BIMS-10). Dans la suite, ces solutions seront vérifiées, à savoir : 1) la version courte de Martin et Sass (2010) (BIMS) ainsi que les modifications statistiques proposées par M*plus* pour améliorer l'ajustement du modèle aux données (BIMS'), 2) la version courte à 12 items avec BM7 et IM7 (BIMS-12) obtenue par l'application stricte de leur procédure ainsi que les améliorations statistiques suggérées par M*plus* (BIMS-12'), et 3) la version courte de 10 items (BIMS-10), cinq par facteur, ainsi que les améliorations suggérées pas M*plus* (BIMS-10').

Les analyses factorielles confirmatoires aboutissent à des indicateurs de qualité enthousiasmants. Ainsi, la BIMS obtient un ajustement satisfaisant entre données et modèle ($\chi^2_{(53)}$ = 170,496, p < 0,001 ; CFI = 0,937 ; TLI = 0,921 ; RMSEA = 0,081 ; WRMR = 1,094). M*plus* propose d'associer, d'une part, les termes d'erreurs de BM9, BM11 avec BM12 ainsi que de BM1 avec BM2 et, d'autre part, d'IM2 avec IM5. Ces corrélations sont conceptuellement acceptables. La liaison entre IM2 et IM5 traite de coopération entre élèves, la relation entre BM1 et BM2 renvoie à des interventions lors de perturbations de la part des élèves, tandis que le lien entre BM9, BM11 et BM12 concerne les règles de fonctionnement en classe. Par ailleurs, elles améliorent significativement l'adéquation entre modèle théorique et données (DIFFTEST : $\chi^2_{(5)}$ = 67,445, p < 0,001). Par la prise en compte de ces améliorations, les valeurs des indicateurs de qualité montrent que l'on obtient un modèle BIMS' (voir figure 1) dont l'adéquation aux données est bonne, voire excellente ($\chi^2_{(48)}$ = 107,840, p <

0,001 ; CFI = 0,968 ; TLI = 0,956 ; RMSEA = 0,061 ; WRMR = 0,841). Les corrélations interfacteurs sont plutôt modérées avec des valeurs de -0,349 et -0,395 pour les deux versions testées dans ce paragraphe.

Tableau 2

Saturations factorielles pour les analyses factorielles exploratoires

	Version longue		**Version courte (Martin & Sass, 2010)**	
	BM	**IM**	**BM**	**IM**
BM1c	0,587*	-0,081	0,574*	-0,133*
BM2c	0,627*	0,062	0,589*	-0,002
BM8c	0,397*	-0,033	0,314*	-0,083
BM9c	0,733*	-0,110*	0,727*	-0,086*
BM11c	0,803*	0,062	0,844*	0,112*
BM12c	0,523*	-0,091*	0,488*	-0,153*
IM2c	-0,066	0,646*	0,006	0,665*
IM3c	-0,216*	0,345*	-0,166*	0,379*
IM5c	0,031	0,789*	0,088*	0,788*
IM6c	-0,057	0,670*	-0,029	0,628*
IM9c	-0,198*	0,503*	-0,182*	0,512*
IM12c	-0,131*	0,683*	-0,080*	0,757*
BM3	0,162*	-0,210*		
BM4	0,243*	0,096		
BM5	0,054	0,769*		
BM6	0,433*	0,424*		
BM7	0,658*	-0,184*		
BM10	0,517*	0,023		
IM1	0,322*	0,236*		
IM4	0,433*	-0,407*		
IM7	0,001	0,788*		
IM8	0,506*	-0,027		
IM10	0,307*	0,073		
IM11	0,286*	0,194*		

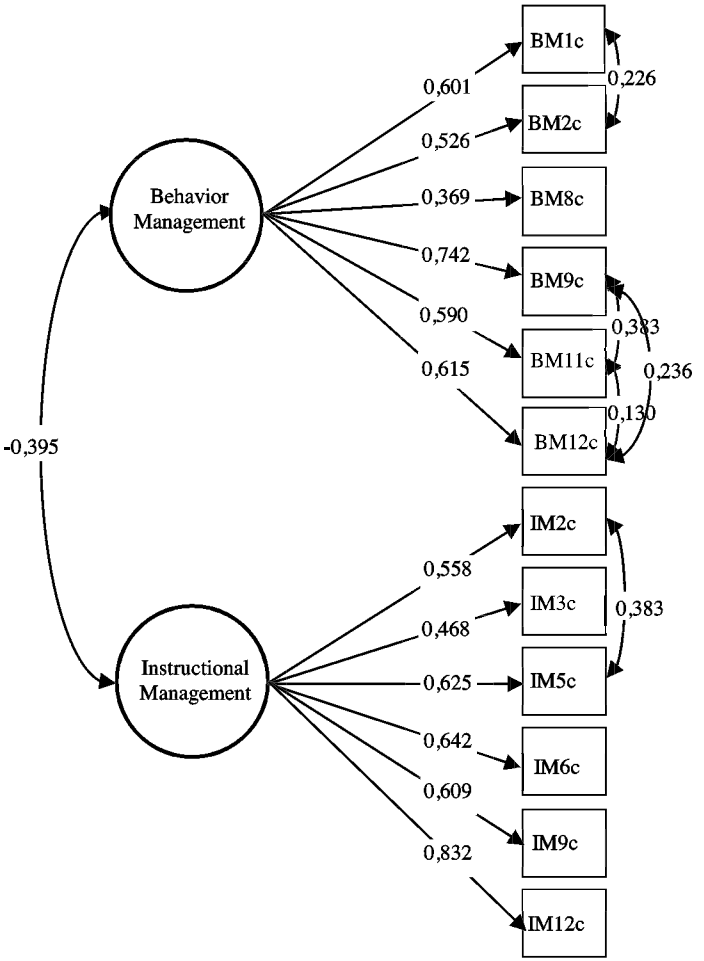


Figure 1. ***Modèle obtenu à partir de la version courte de Martin et Sass (2010) avec les améliorations suggérées par M*plus – ***BIMS'***

Comme il a été précisé ci-dessus, avec une observation des saturations selon les critères de Martin et Sass (2010), il y a deux solutions. Pour ce qui est de la BIMS-12, solution où l'on conserve 12 items mais où l'on remplace BM8 par BM7 et IM3 par IM7, un ajustement satisfaisant est obtenu ($\chi^2_{(53)}$ = 159,492, p < 0,001 ; CFI = 0,947 ; TLI = 0,934 ; RMSEA = 0,089 ; WRMR = 1,163). M*plus* suggère de corréler les termes d'erreurs de BM1 et BM2, ainsi que d'IM2 et IM5. Étant donné que ces deux associations sont sémantiquement pertinentes, comme il a été expliqué

plus haut, il en ressort le modèle BIMS-12', présenté en figure 2 et ayant des indicateurs de qualité intéressants ($\chi^2_{(51)}$ = 164,416, p < 0,001 ; CFI = 0,958 ; TLI = 0,946 ; RMSEA = 0,081 ; WRMR = 1,061), mais l'amélioration statistique par rapport au modèle sans association est faible quoique significative (DIFFTEST : $\chi^2_{(2)}$ = 8,316, p = 0,015). Les corrélations inter-facteurs sont de -0,314 pour la BIMS-12 et de -0,324 pour la BIMS-12'.

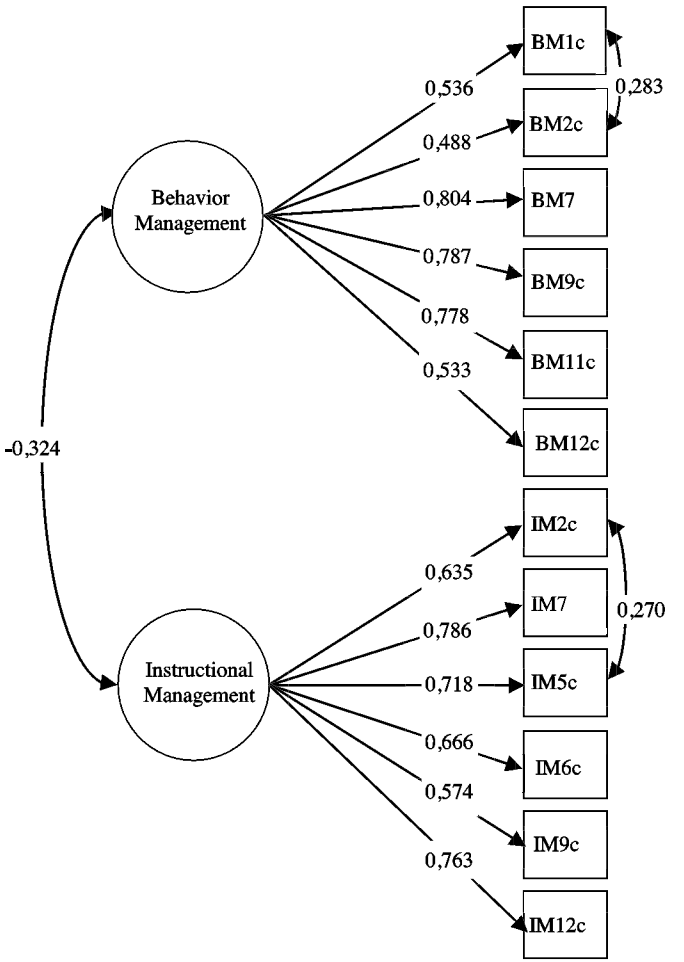


Figure 2. ***Modèle obtenu à partir de l'application de la procédure de Martin et Sass (2010) avec les améliorations suggérées par Mplus – BIMS-12'***

La suppression de deux items de la BIMS de Martin et Sass (2010), à savoir BM8 et IM3, repose sur le fait que leur saturation est plus petite que 0,40 dans leur facteur théorique et sur l'envie de ne pas ajouter des items étrangers à la solution courte retenue par Martin et Sass (2010). Ce modèle, la BIMS-10, obtient d'intéressants indicateurs de qualité d'ajustement ($\chi^2_{(34)}$ = 141,774, p < 0,001 ; CFI = 0,937 ; TLI = 0,916 ; RMSEA = 0,097 ; WRMR = 1,130). M*plus* suggère à nouveau des modifications en termes de corrélation entre scores d'erreurs, mais seules certaines ont un sens théorique : BM1 et BM2 traitent de la limitation des bavardages, tandis que BM11 et BM9 concernent l'assurance du respect constant et permanent des règles. Le programme suggère l'association des termes d'erreurs d'IM9 et de BM9, comme c'est le cas chez Martin et Sass (2010). Cependant, cette proposition a été ignorée, car, comme l'avouent les auteurs, elle améliore certes l'ajustement statistique de leur modèle, mais n'a pas de signification conceptuelle. L'analyse des données de cette étude conduit au modèle BIMS-10', suggéré en figure 3. Il obtient un bon ajustement ($\chi^2_{(32)}$ = 115,906, p < 0,001 ; CFI = 0,951 ; TLI = 0,931 ; RMSEA = 0,088 ; WRMR = 1,011). L'amélioration ainsi apportée est statistiquement significative (DIFFTEST : $\chi^2_{(2)}$ = 30,746, p < 0,001). Les corrélations interfacteurs sont de -0,314 et -0,343 pour ces deux versions.

Concernant les coefficients de consistance interne, la BIMS obtient un alpha de Cronbach de 0,717 pour la sous-échelle BM et de 0,751 pour l'IM. Ces valeurs passent respectivement à 0,727 et 0,755 pour la version BIMS-10 et à 0,782 et 0,805 pour la version BIMS-12. Le calcul du coefficient ajusté de consistance interne proposé par Brown (2006) mène aux valeurs suivantes pour la BIMS : pour BM, 0,780 sans corrélation entre termes d'erreurs (BIMS) et 0,668 en prenant en compte ces corrélations (BIMS') ; pour IM, ces valeurs sont respectivement de 0,808 et 0,761. La BIMS-12 obtient les valeurs suivantes : pour BM, 0,833 et 0,798 avec les corrélations entre termes d'erreurs ; pour IM, 0,852 et 0,824 avec les corrélations. La BIMS-10 permet d'obtenir des valeurs de consistance interne ajustée de 0,793, ou 0,688 avec les corrélations entre termes d'erreurs pour BM ; pour IM, à une valeur unique de 0,812, puisque M*plus* ne suggère pas de corrélation entre termes d'erreurs pour cette sous-échelle. Avec des coefficients oscillant autour de 0,70 et approchant 0,90 dans certains cas, nous pouvons estimer que les sous-échelles obtenues par le biais de cet échantillon d'enseignants en formation possèdent une bonne consistance interne.

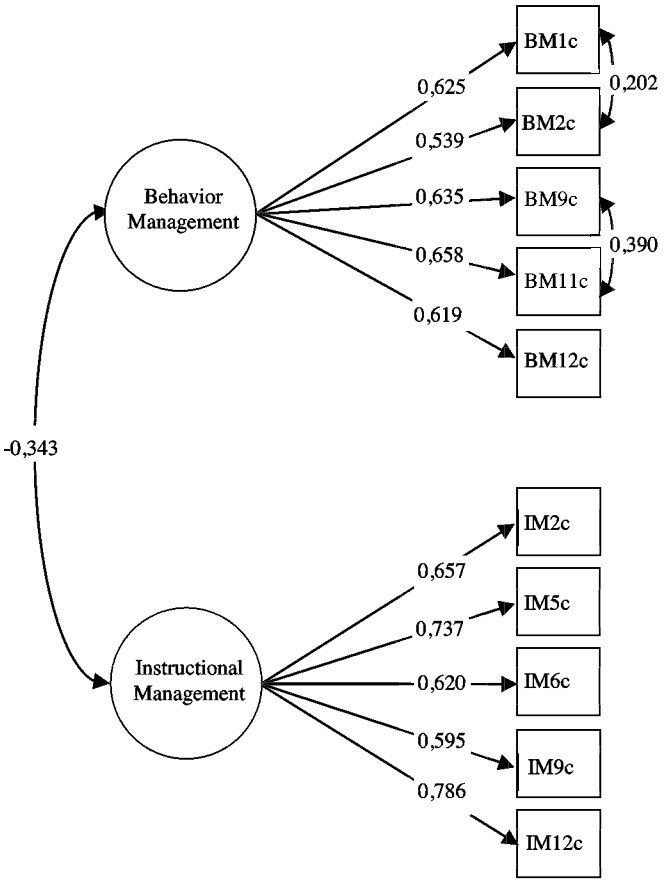


Figure 3. ***Modèle obtenu par la suppression de BM8 et IM3 dans Martin et Sass (2010) avec les améliorations suggérées par M*plus – *BIMS-10'***

Discussion

Martin et Sass (2010) suggèrent aux chercheurs de collecter des données à partir de leurs 24 items afin de vérifier si la même version raccourcie de 12 items se manifeste ailleurs. Ils estiment que leur instrument, qui se trouve d'après eux dans une phase de développement peu avancée, présente des propriétés psychométriques prometteuses qu'il conviendrait de vérifier à l'aide d'autres échantillons. Dans la présente étude, leurs 24 items ont donc été proposés à un échantillon de 338 enseignants en formation dans le but d'appliquer à ces données les mêmes

traitements statistiques que les chercheurs. Dans l'ensemble, la version courte proposée par Martin et Sass (2010) tient relativement bien la route. Néanmoins, il apparaît que d'autres versions se défendent tout autant. Les propriétés psychométriques obtenues rejoignent celles que Martin et Sass (2010) révèlent dans leur article, voire les dépassent dans certains cas.

Pourtant, les données de cette étude impliquent deux différences avec la version de Martin et Sass (2010). Premièrement, les améliorations suggérées par M*plus* ne sont pas conformes à celles que les auteurs obtiennent. Dans leur cas, l'amélioration statistique apportée par la corrélation des termes d'erreurs de BM9 et IM9 n'a pas de cohérence théorique. Dans la présente étude, les corrélations apportent une amélioration statistique qui a une cohérence conceptuelle. Deuxièmement, l'application à la lettre de leur procédure de sélection des items mène à des constats légèrement différents que ceux retenus par Martin et Sass (2010). Les solutions proposées dans cette contribution consistent à intégrer une autre paire d'items en remplacement d'une paire d'items qui ne sature pas suffisamment dans son facteur, ou à éliminer ces deux items. Ces analyses conduisent au constat que la version de Martin et Sass (2010) appliquée à l'échantillon de cette étude avec les améliorations suggérées par M*plus* obtient les meilleures qualités psychométriques, selon les résultats de l'analyse factorielle confirmatoire. Toutefois, les indices de consistance interne ne sont pas les meilleurs pour cette version.

Le développement d'outil de mesure des conduites pédagogiques déclarées des enseignants est important pour étudier les dynamiques de classe et la stabilité ou l'instabilité de l'enseignement. L'outil de Martin et Sass (2010) présente l'avantage d'être relativement court avec ses 12 items. Néanmoins, d'autres recherches devraient être menées avec les 24 items pour vérifier si les mêmes 12 items sont conservés dans d'autres contextes, comme le suggèrent Martin et Sass (2010). Cela a été le cas avec l'échantillon de cette étude, mais des comparaisons internationales avec de plus grands échantillons pourraient asseoir la pertinence des items de la BIMS pour mesurer les pratiques déclarées des enseignants sur la gestion de la classe, et des recherches pourraient être proposées dans ce sens.

D'autres voies de recherche sont envisageables. Ainsi, Martin et Sass (2010) suggèrent notamment de vérifier la qualité de leur outil en variant l'ordre des items pour diminuer la potentielle influence interitems. Il serait aussi intéressant d'entreprendre des études pour vérifier la concordance

entre les pratiques déclarées par le biais de la BIMS et celles qui sont observées en classe. En effet, les recherches étudiant le lien entre croyances et pratiques sont peu nombreuses (Fang, 1996), mais nécessaires. Cette voie permettrait notamment de valider la BIMS d'une autre manière. Aussi, il faudrait vérifier dans quelle mesure les pratiques déclarées par le biais de la BIMS correspondent à ce qui se passe effectivement en classe. De surcroît, la BIMS pourrait être utilisée dans des études transversales ou longitudinales afin de relever la manière dont se développent les pratiques ou croyances des enseignants en formation et en fonction, en matière de gestion de la classe.

Bien entendu, une recherche comme celle présentée ici a des limites importantes. La première concerne incontestablement la taille de l'échantillon. Alors que Martin et Sass (2010) ont bénéficié de la participation de 550 enseignants chevronnés dont la moitié intervenait à l'élémentaire et l'autre moitié au secondaire, l'échantillon de cette étude est composé de 338 enseignants en formation, avec une proportion identique entre les deux ordres d'enseignement. Le sondage d'un plus grand nombre de participants et des analyses des différences de structures factorielles des pratiques managériales déclarées en fonction des ordres d'enseignement ou d'autres variables seraient importants. Martin et Sass (2010) suggèrent qu'il est probable que l'on n'obtienne pas le même modèle à partir de données portant sur des enseignants du primaire ou sur des enseignants du secondaire. L'échantillon de cette étude, tout comme celui de Martin et Sass (2010) d'ailleurs, aurait permis cette analyse. Toutefois, vu le nombre restreint de participants, il a été décidé, comme l'ont fait Martin et Sass (2010), de ne pas procéder à cette analyse ici. D'autres recherches pourraient entreprendre cette comparaison.

L'analyse du χ^2 comme indicateur de qualité d'ajustement des modèles pourrait pousser certains à rejeter la BIMS. Pour ces personnes, il conviendrait peut-être d'adapter la BIMS au contexte pédagogique genevois, voire de créer un outil adapté au système de ce canton de la Suisse. Il est vrai que pareille entreprise est intéressante et que des recherches pourraient être menées dans ce sens. Néanmoins, deux arguments pourraient être avancés ici afin d'au moins intégrer dans un premier temps la BIMS dans cet instrument adapté, avant de l'écarter définitivement. Premièrement, il faut rappeler la critique que Brown (2006) formule sur la place historique du χ^2 dans l'interprétation de l'ajustement

des modèles et sur son inexactitude en cas d'échantillon dépassant 200 individus. C'est le cas de l'échantillon de cette étude et aussi de l'échantillon de Martin et Sass (2010). Deuxièmement, il faut mentionner qu'il serait problématique d'utiliser des outils de mesure différents pour une même variable si des comparaisons internationales étaient visées. D'où l'intérêt de vérifier la qualité de la BIMS dans d'autres contextes, d'autant plus qu'en milieu genevois avec des enseignants en formation, cette analyse mène à des constats remarquablement similaires à ceux de Martin et Sass (2010).

Enfin, l'examen du développement de la BIMS montre que le score de management des individus a évolué au fil du temps. La BIMS est passée d'une hypothèse et du calcul d'un continuum unidimensionnel opposant interventionnisme et non-interventionnisme à la génération de deux scores correspondant chacun à une dimension donnée de la gestion de la classe. Des recherches devraient être entreprises pour vérifier si les pratiques managériales déclarées des enseignants sont unidimensionnelles ou bidimensionnelles. Les corrélations interfacteurs raisonnables obtenues dans cette étude semblent privilégier l'hypothèse de la bidimensionnalité. Cette analyse est relativement simple à effectuer, puisqu'il suffit de fixer à 1, dans M*plus*, la corrélation entre les deux dimensions de la BIMS. M*plus* refusant de converger dans tous les cas, ces analyses n'ont pas été rapportées dans cet article. Ceci renforce d'une certaine manière l'hypothèse selon laquelle les pratiques managériales déclarées seraient bidimensionnelles. Des analyses reposant sur un échantillon plus important sont nécessaires pour soutenir cette affirmation. Néanmoins, cette voie serait prometteuse, car elle permettrait de faire l'hypothèse de l'existence d'au moins quatre profils pédagogiques dont la recherche pourrait analyser les effets sur les performances et/ou sur la motivation des élèves, voire leur utilisation variable par les enseignants en fonction des situations d'enseignement ou des profils des élèves.

On l'aura compris, les limites de la BIMS sont nombreuses, mais ses atouts le sont tout autant. Reposant sur des qualités psychométriques fortes, elle possède un potentiel de recherche important. Nous espérons que cet article donnera envie à des chercheurs de l'utiliser et d'en vérifier la pertinence, par exemple en suivant les quelques pistes suggérées.

Réception : 12 juillet 2013
Version finale : 28 janvier 2015
Acceptation : 16 février 2015

NOTES

1. CFI = *comparative fit index*; TLI = *Tucker-Lewis index*; RMSEA = *root mean square error of approximation*; SRMR = *standardized root mean square residual*; WRMR = *weighted root mean square residual*.
2. www.statmodel.com (voir le forum de discussion M*plus* Discussion).
3. En termes mathématiques, l'expression de la formule est la suivante :

$$\text{Fidélité ajustée} = \frac{(\Sigma\lambda_i)^2}{(\Sigma\lambda_i)^2 + \Sigma\theta_{ii} + 2\Sigma\theta_{ij}}$$

RÉFÉRENCES

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: The Guilford Press.

Browne, M. W., & Cudek, R. (1993). Alternate ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research*, *38*(1), 47-65. doi: http://dx.doi.org/10.1080/0013188960380104

Furr, M. (2011). *Scale construction and psychometrics for social and personality psychology*. London, UK: Sage. doi: http://dx.doi.org/10.4135/9781446287866

Glickman, C. D., & Tamashiro, R. T. (1980). Clarifying teachers' beliefs about discipline. *Educational Leadership*, *37*, 459-464.

Henson, R. K. (2003). Relationships between preservice teachers' self-efficacy, task analysis, and classroom management beliefs. *Research in the Schools*, *10*(1), 53-62.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55. doi: http://dx.doi.org/10.1080/10705519909540118

Laveault, D. (2012). Soixante ans de bon et mauvais usages du alpha de Cronbach. *Mesure et évaluation en éducation*, *35*(2), 1-7. doi: http://dx.doi.org/10.7202/1024716ar

Martin, N. K., & Baldwin, B. (1993a, April). *Validation of an Inventory of Classroom Management Style: Differences between novice and experienced teachers*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Martin, N. K., & Baldwin, B. (1993b, November). *An examination of the construct validity of the Inventory of Classroom Management Style*. Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans, LA.

Martin, N. K., & Baldwin, B. (1994, January). *Beliefs regarding Classroom Management Style: Differences between novice and experienced teachers*. Paper presented at the annual conference of the Southwest Educational Research Association, San Antonio, TX.

Martin, N. K., & Sass, D. A. (2010). Construct validation of the Behavior and Instructional Management Scale. *Teaching and Teacher Education*, *26*, 1124-1135. doi: http://dx.doi.org/10.1016/j.tate.2009.12.001

Martin, N. K., & Yin, Z. (1997, January). *Attitudes and Beliefs regarding Classroom Management Style: Differences between male and female teachers*. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX.

Martin, N. K., Yin, Z., & Baldwin, B. (1998a, April). *Classroom management training, class size and graduate study: Do these variables impact teachers' beliefs regarding classroom management style?* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Martin, N. K., Yin, Z., & Baldwin, B. (1998b). Construct validation of the Attitudes and Beliefs on Classroom Control Inventory. *Journal of Classroom Interaction*, *33*(2), 6-15.

Martin, N. K., Yin, Z., & Mayall, H. (2007). The Attitudes and Beliefs on Classroom Control Inventory revised and revisited: A continuation of construct validation. *Journal of Classroom Interaction*, *42*(2), 11-20.

Muthén, L. K., & Muthén, B. O. (2002). *M*plus *statistical analysis with latent variables: User's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Ritter, J. T., & Hancock, D. R. (2007). Exploring the relationship between certification sources, experience levels, and classroom management orientations of classroom teachers. *Teaching and Teacher Education*, *23*, 1206-1216. doi: http://dx.doi.org/10.1016/j.tate.2006.04.013

Savran, A., & Cakiroglu, J. (2003). Differences between elementary and secondary preservice teachers' perceived efficacy beliefs and their classroom management beliefs. *The Turkish Online Journal of Educational Technology*, *2*(4), 15-20.

Sokal, L., Smith, D. G., & Mowat, H. (2003). Alternative certification teachers' attitudes toward classroom management. *The High School Journal*, *86*(3), 8-16. doi: http://dx.doi.org/10.1353/hsj.2003.0004

Tschannen-Moran, M., & Woolfolk-Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*, 783-805. doi: http://dx.doi.org/10.1016/s0742-051x(01)00036-1

Wang, J., & Wang, X. (2012). Structural equation modeling: Applications using M*plus*. Chichester, UK: John Wiley & Sons. doi: http://dx.doi.org/10.1002/9781118356258.ch1

Wolfgang, C. H., & Glickman, C. D. (1986). *Solving discipline problems: Strategies for classroom teachers* (2nd ed.). Newton, MA: Allyn & Bacon.

Yilmaz, H., & Çavaş, P. H. (2008). The effect of the teaching practice on pre-service elementary teachers' science teaching efficacy and classroom management beliefs. *Eurasia Journal of Mathematics, Science and Technology Education*, *4*(1), 45-54.

Annexe

Version française de la BIMS

Numéro de l'item	Affirmation
BM1c	J'interviens presque toujours quand des élèves parlent à des moments inappropriés durant l'enseignement en classe.
IM1	J'utilise l'instruction collective de la classe pour assurer un enseignement structuré.
BM2c	Je limite fortement les bavardages des élèves dans la salle de classe.
*IM2c	J'utilise presque toujours l'apprentissage coopératif afin d'explorer des sujets/contenus en classe.
BM3	Je récompense les élèves pour les bons comportements en classe.
*IM3c	Je favorise l'engagement des élèves dans des discussions actives concernant des éléments ayant une application dans le monde réel.
BM4	Si un élève parle à un voisin, je l'éloigne des autres élèves.
IM4	J'établis et je maintiens une routine d'enseignement quotidienne dans ma classe.
*BM5	J'utilise les suggestions des élèves pour créer des règles de vie en classe.
*IM5c	J'utilise presque toujours le travail en groupes dans mon enseignement.
*BM6	J'autorise les élèves à se lever de leur chaise sans qu'ils en aient demandé la permission.
*IM6c	J'utilise les apports des élèves lorsque nous créons des projets pour eux.
BM7	Je suis strict lorsqu'il est question de respect des règlements et de conformité aux règles en classe.
*IM7	J'utilise presque toujours l'apprentissage par enquête ou découverte dans ma classe.
BM8c	Je redirige fermement l'attention des élèves vers la matière lorsqu'ils sont en dehors de l'activité.
IM8	Je gère les transitions des élèves d'une activité d'apprentissage vers une autre.
BM9c	J'insiste pour que, dans ma classe, les élèves respectent les règles en tout temps.
*IM9c	J'ajuste presque toujours mon enseignement afin de répondre aux besoins individuels des élèves.
BM10	Je gère et surveille minutieusement les activités hors tâche durant l'enseignement.
IM10	J'utilise presque toujours l'instruction directe/frontale en groupe collectif lorsque j'enseigne.
BM11c	Je mets strictement en application les règles de conduite en classe pour contrôler les comportements des élèves.
IM11	Je ne dévie pas de ma planification du cours (plans d'activités d'apprentissage).
BM12c	Si le comportement des élèves est provocateur, je leur demande de se conformer à mes règles de conduite en classe.
*IM12c	J'utilise presque toujours une approche pédagogique qui encourage l'interaction et la communication entre les élèves.

c = version courte de Martin et Sass (2010)
* = item au score inversé.