Document generated on 05/09/2025 3:42 a.m.

Mesure et évaluation en éducation

La catégorisation des élèves par les enseignants : une étude critique des propositions de Hofer

Philippe Wanlin, Marie-Louise Aliprandi, Angela Mossaz and Malika Revilloud

Volume 39, Number 1, 2016

URI: https://id.erudit.org/iderudit/1036706ar
DOI: https://doi.org/10.7202/1036706ar

See table of contents

Publisher(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (print)
2368-2000 (digital)

Explore this journal

Cite this article

Wanlin, P., Aliprandi, M.-L., Mossaz, A. & Revilloud, M. (2016). La catégorisation des élèves par les enseignants : une étude critique des propositions de Hofer. *Mesure et évaluation en éducation*, *39*(1), 67–94. https://doi.org/10.7202/1036706ar

Article abstract

This article presents Hofer's (1981, 1986) research about students' categories in teachers' cognition. Its aim is to extract Hofers' methodology and to supply it with approaches used by researchers interested in the categorization of persons or objects (Sternberg, 2007; Reed, 2011). Nine Geneva elementary school female teachers took part in this research that pursue two goals: (a) verifying the relevance of a statistical method (cluster analysis) to identify students' categories in teachers' general pedagogical knowledge, and (b) confirm whether or not the central-peripheral internal structure of the conclusions Hofer thought he was demonstrating is the result of his statistical algorithms. Results are discussed in terms of methodological and empirical perspectives.

Tous droits réservés © ADMEE-Canada - Université Laval, 2016

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

érudit

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

La catégorisation des élèves par les enseignants : une étude critique des propositions de Hofer

Philippe Wanlin
Université de Genève (IUFE)

Marie-Louise Aliprandi

Angela Mossaz

Malika Revilloud
Université de Genève (IUFE-FPSE)

Mots-clés : connaissances, évaluation, enseignants, élèves

Notre contribution présente les recherches de Hofer (1981, 1986) qui se sont intéressées aux catégories d'élèves contenues dans le répertoire cognitif des enseignants. Elle se penche plus spécifiquement sur l'approche utilisée par cet auteur afin d'en extraire la méthodologie, dont nous avons fait usage notamment à l'aide des approches utilisées dans les recherches sur la catégorisation des objets et des personnes (Sternberg, 2007 ; Reed, 2011). Neuf enseignantes genevoises du primaire ont participé à notre dispositif de recherche, qui poursuit deux objectifs : (a) vérifier la pertinence du recours à une méthode statistique, soit l'analyse par grappes (cluster analysis), pour identifier les catégories d'élèves disponibles dans le bagage cognitif des enseignants, et (b) examiner si la structure interne centrale-périphérique de catégories isolées de connaissances que Hofer pense démontrer n'est pas le produit des algorithmes statistiques qu'il utilise. Nous concluons par une discussion sur nos résultats concernant la méthodologie et les perspectives de recherche.

Keywords: knowledge, evaluation, teachers, students

This article presents Hofer's (1981, 1986) research about students' categories in teachers' cognition. Its aim is to extract Hofers' methodology and to supply it with approaches used by researchers interested in the categorization of persons or objects (Sternberg, 2007; Reed, 2011). Nine Geneva elementary school female teachers took part in this research that pursue two goals: (a) verifying the relevance of a statistical method (cluster analysis) to identify students' categories in teachers' general pedagogical knowledge, and (b) confirm whether or not the central-peripheral internal structure of the conclusions Hofer thought he was demonstrating is the result of his statistical algorithms. Results are discussed in terms of methodological and empirical perspectives.

Palavras-chave: conhecimentos, avaliação, professores, alunos

Este artigo apresenta as investigações de Hofer (1981, 1986) sobre as categorias de alunos incluídos no repertório cognitivo dos professores. Mais especificamente, analisa a abordagem utilizada por este autor para extrair a metodologia que usamos, nomeadamente no apoio às abordagens utilizadas nas investigações sobre a categorização de objetos e pessoas (Sternberg, 2007; Reed, 2011). Nove professoras do ensino básico (1.º ciclo) de Genebra participaram no nosso dispositivo de investigação, o qual tem dois objetivos: (1) verificar a pertinência do recurso a um método estatístico – análise de conglomerados (cluster analysis) – para identificar as categorias de alunos disponíveis na bagagem cognitiva dos professores, e (2) analisar se a estrutura interna central-periférica de categorias isoladas de conhecimentos que Hofer pensa demonstrar não é o produto de algoritmos estatísticos que ele utiliza. Concluímos com uma discussão dos nossos resultados no que respeita à metodologia e às perspetivas de investigação.

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

À l'Université de Genève, l'IUFE est l'Institut universitaire de formation des enseignants, tandis que la FPSE est la Faculté de psychologie et des sciences de l'éducation.

Introduction

Plusieurs recherches ont identifié des catégories d'élèves dans le répertoire cognitif des enseignants (Thelen, 1967 ; Hörstermann, Krolak-Schwerdt & Fischbach, 2010), mais nous ne nous intéresserons dans cet article qu'à celles menées par Hofer (1981, 1986). Ce chercheur exploite des données de questionnaires de jugements qu'il soumet à des analyses par grappes (*cluster*s) pour générer des profils d'élèves. Il observe que les enseignants catégorisent leurs élèves en sous-ensembles. En outre, Hofer pense prouver que les représentations des enseignants sur les élèves sont compartimentées en grappes dont la structure est composée d'une périphérie gravitant autour d'un noyau central. Cependant, sa méthodologie est remise en question par certaines recherches, qui considèrent que l'organisation interne des catégories est un artifice statistique.

Nous relatons, dans cet article, une recherche inspirée de l'approche de Hofer. Neuf enseignantes du primaire du canton de Genève ont évalué tous les élèves de leur classe avec une échelle de jugements. Ces jugements ont été soumis à une analyse par grappes pour reproduire une catégorisation à la manière de Hofer. Pour vérifier l'organisation centrale-périphérique, nous leur avons demandé d'effectuer deux tâches supplémentaires consistant à, d'une part, répartir leurs élèves au sein du système de grappes identifié et, d'autre part, à se prononcer quant au degré de typicité de chacun des élèves compte tenu des définitions des grappes. Par ailleurs, nous leur avons demandé de procéder à un regroupement libre de leurs élèves comme l'a utilisé Morine-Dershimer (1978, 1979), puis d'exprimer un degré de typicité de chaque élève à ces groupes. Cette procédure de recherche permet de poursuivre deux objectifs : d'un côté, vérifier la pertinence du recours à l'analyse par grappes pour examiner la catégorisation des élèves par les enseignants et, de l'autre, examiner la pertinence du maintien de l'hypothèse d'une organisation des représentations en des entités isolées organisées de manière centrale-périphérique lors d'une analyse par grappes.

Les travaux de Hofer

Pour Hofer (1981), le chercheur doit identifier des types d'élèves au départ des jugements des enseignants selon une ou plusieurs caractéristiques. Il présente deux recherches allant dans ce sens dans son article : la première porte sur des élèves du secondaire, tandis que la seconde porte sur des élèves du primaire.

Les regroupements d'élèves du secondaire selon les jugements de leurs enseignants

Hofer (1981) analyse des données rassemblées auprès de 15 enseignants allemands qui enseignent leur langue. Ils ont été priés de juger tous les élèves de l'une de leurs classes selon 25 caractéristiques liées à des échelles de type Likert à 7 modalités. Par exemple, pour la caractéristique « attitude que l'élève a face à l'école », l'échelle allait de 1 – attitude très défavorable à 7 – attitude très favorable (voir Tableau 3 pour d'autres exemples de caractéristiques). Hofer a ensuite soumis les matrices de jugements à une analyse par grappes selon deux algorithmes agglomératifs. Cette procédure lui a permis de mettre en évidence une catégorisation des 358 élèves de son échantillon en cinq grappes, qu'il décrit à partir de la comparaison de leurs moyennes aux 25 variables jugées. Ces cinq grappes peuvent être interprétées comme étant deux groupes de bons et deux groupes de mauvais élèves ainsi qu'un groupe d'élèves moyens.

- La première grappe comprend 64 élèves obtenant de bonnes notes que les enseignants estiment intelligents, appliqués, disciplinés, actifs et contrôlables sur le plan du comportement en classe ;
- La deuxième grappe renferme 120 élèves obtenant également de bonnes notes ainsi que de hautes valeurs de talent, de perspicacité et de discipline. Ils font preuve de retenue sociale : ils sont sensibles, calmes, modestes, simples et participent peu ;
- La troisième grappe contient 60 élèves moyens dans les performances scolaires et les autres variables. Ils sont considérés comme étant fermés, peu sûrs d'eux, timides, calmes et distants, mais relativement talentueux, appliqués, disposés à faire des efforts et disciplinés ;
- La quatrième grappe englobe 74 élèves très peu disciplinés avec une intelligence jugée moyenne, une haute activité sociale et plutôt une mauvaise relation avec le travail. Les trublions et les dérangeurs de tous les niveaux de rendement sont contenus dans cette grappe ;

- Les 40 élèves de la cinquième grappe sont perçus comme étant peu intéressés, peu doués, sans ambition ni perspicacité au travail et manquant d'activité sociale.

Les regroupements d'élèves du primaire selon les jugements de leurs enseignants

Pour vérifier si la même typologie se retrouve dans d'autres niveaux scolaires, Hofer (1981) a mené la même étude auprès de 167 élèves de cinq classes de deuxième année du primaire. Pour cet échantillon, le chercheur procède à une mise en grappes (*clustering*) initiale et à une mise en grappes adaptée. Pour sa mise en grappes initiale, Hofer utilise deux méthodes agglomératives hiérarchiques conduisant à une classification totalement identique des élèves : une solution à quatre ensembles. Hofer ne décrit pas les quatre ensembles d'élèves du primaire dans son article, mais précise que, comparativement aux grappes du secondaire, les deux premiers groupes fusionnent et que les autres sont inchangés (Hofer, 1981). En d'autres termes, il obtient une grappe de bons élèves, deux d'élèves faibles et une grappe d'élèves moyens.

Pour comparer ces données avec celles du secondaire, Hofer procède à une nouvelle catégorisation de ces 167 élèves du primaire en forçant une solution à cinq grappes, puis calcule leurs scores moyens aux 25 variables (voir Tableau 3 pour des exemples). Il corrèle ensuite les scores moyens des grappes du primaire avec ceux des grappes du secondaire. Il obtient des corrélations allant de 0,78 à 0,98. Hofer en conclut qu'il y a une haute superposition dans la manière dont des enseignants de différents niveaux s'y prennent pour grouper les élèves selon leur ressemblance.

La tendance au regroupement comme artifice statistique

Hofer déduit des deux études qui précèdent que les enseignants résument cognitivement les élèves en un système de types implicites comportant de quatre à cinq profils. Pour lui, la méthode d'analyse par grappes est valide pour les identifier. Cette hypothèse est cependant critiquée par Oldenbürger (1986) et Friedrich (1979). Avec les données de Hofer (1969) portant sur 36 enseignants et leurs 810 élèves, Oldenbürger (1986) a testé cette hypothèse. Il écrit qu'avant de pouvoir soumettre les matrices de données à une analyse par grappes, ayant par essence la propriété de structurer les données en ensembles hétérogènes d'éléments homogènes (Everitt, Landau, Leese & Stahl, 2011), il faut analyser leur structure interne à

l'aide d'algorithmes statistiques adaptés pour s'assurer qu'elles contiennent des ensembles différents d'éléments ressemblants. Les algorithmes statistiques qu'il présente dans son article sont d'une extrême complexité, c'est pourquoi nous ne rapporterons que sa conclusion. D'après ses résultats, les matrices de données de Hofer (1969) peuvent être considérées comme contenant des ensembles distincts d'élèves. Néanmoins, Oldenbürger (1986) montre que l'analyse par grappes n'est pas en mesure de les identifier. Il estime, de ce fait, que les classifications de Hofer sont des artefacts de recherche.

Friedrich (1979) émet une critique semblable à l'aide des jugements de 76 enseignants au secondaire autrichiens : il n'identifie pas des ensembles homogènes d'élèves au sein de sa matrice de données. Il montre aussi que l'échelle de jugements de Hofer contient des dimensions que les enseignants n'utilisent pas. De plus, il constate que les items à pôles opposés de Hofer n'entrent pas en résonance avec les représentations et critères des enseignants. Par exemple, dans l'échelle de Hofer, l'opposé d'intelligent est « bête », alors que Friedrich obtient que, pour les enseignants, le contraire d'intelligent est « a davantage besoin d'aide pour atteindre des objectifs pédagogiques ».

Hofer n'a pas ignoré ces critiques. Pour vérifier la pertinence de sa méthode, Hofer et Köpke (1987) ont comparé l'analyse par grappes à l'approche de Morine-Dershimer (1978, 1979). Cette dernière demande aux enseignants de regrouper librement leurs élèves selon leur ressemblance (pour une description plus complète, voir la section Méthode). Hofer et Köpke (1987) ont demandé à une vingtaine d'enseignants de cinq écoles de juger leurs élèves avec l'échelle de jugements de Hofer (1981), puis de les regrouper librement selon la méthode de Morine-Dershimer (1978). Dans cette étude, lors d'un entretien, les enseignants groupaient d'abord leurs élèves selon leurs similarités en autant de groupes qu'ils le souhaitaient, puis ils devaient expliquer les critères utilisés. Les analyses enseignant-par-enseignant de Hofer et Köpke (1987) montrent des taux de recouvrement décevants entre les catégories obtenues par regroupement libre et celles obtenues par mise en grappes des scores de jugements. Ils écrivent que, même si les partitionnements obtenus par mise en grappes ne sont pas le fruit du hasard, les coefficients de concordance avec les regroupements libres ne sont pas suffisamment élevés et que seulement les deux tiers des élèves obtiennent une catégorisation similaire par les deux méthodes. Ils concluent que les méthodes par mise en grappes de juge-

ments et par regroupement libre ne sont pas parfaitement adéquates pour déceler les catégories imperméables d'élèves semblables contenues dans le bagage cognitif des enseignants. En effet, ils estiment qu'un certain nombre d'élèves obtiennent des catégorisations chevauchantes sur plusieurs ensembles, mais les algorithmes statistiques utilisés ne permettent pas de le prendre en compte.

La structure interne des grappes d'élèves

Hofer (1986) interprète ses résultats précédents sur l'organisation interne des représentations des enseignants en se référant à la théorie des prototypes d'Eleanor Rosch (1973, 1975, 1978). Il estime que les représentations sont compartimentées en grappes distinctes qui adoptent une organisation de type central-périphérique. Cette hypothèse entre en concordance avec la théorie de Green (1971). Pour lui, les croyances adoptent une structure centrale-périphérique à l'intérieur de grappes de croyances qui n'ont généralement pas de liens entre elles (leurs frontières sont fermées). Toutefois, Green n'exclut pas la possibilité que les frontières entre ces grappes soient poreuses pour certains éléments cognitifs.

Pour les théories conjecturant une structure centrale-périphérique, les représentations contiennent des éléments regroupés en catégories selon leur proximité ou leur typicalité. Ces ensembles seraient formés à partir de traits caractéristiques en nombre variable décrivant un ou des modèles prototypiques de la catégorie. Ce ou ces modèles seraient représentés soit par un ou des « centroïdes », à savoir une distance moyenne « virtuelle » de la catégorie, soit par un ou plusieurs de ses membres considérés comme les meilleurs exemplaires (Hofer, 1986 ; Medin, 2004 ; Reed, 2011 ; Rosch, 1978 ; Sternberg, 2007).

Hofer (1986) pense que ses recherches ont démontré que les représentations des enseignants sur leurs élèves adoptent cette structure. Il serait possible, selon lui, de montrer, pour la mise en grappes de jugements, que les représentations des enseignants sur leurs élèves sont composées d'un nombre restreint de catégories, que ces catégories sont fermées et que les élèves qui les composent sont proches les uns des autres et éloignés des élèves des autres catégories. Cependant, en utilisant l'analyse par grappes, Hofer a obligé ses données à adopter cette structure. Aussi, nous pensons que sa recherche devrait être poursuivie par trois stratégies de recueil de données pour répondre à trois questions de recherche.

Premièrement, et afin de vérifier la pertinence de l'approche par mise en grappes de jugements, il faudrait demander à des enseignants de catégoriser leurs élèves à l'intérieur du système de grappes obtenues par l'algorithme statistique. Notre première question de recherche est donc : les enseignants assignent-ils à leurs élèves les mêmes étiquettes de grappes que celles que l'algorithme statistique leur confère à partir de leurs jugements ?

La deuxième question de recherche interroge la pertinence de l'hypothèse de l'organisation centrale-périphérique des élèves en grappes isolées les unes des autres. Ici, nous adoptons une approche inspirée des recherches classiques en matière de catégorisation d'objets (Reed, 2011 ; Sternberg, 2007) qui analyse communément dans quelle mesure des sujets estiment que des objets sont proches des définitions des catégories. Nous demanderons donc aux enseignantes d'estimer la distance de chacun de leurs élèves à la catégorie à laquelle elles les ont assignés ainsi qu'aux autres catégories.

Troisièmement, conformément à la méthode de Morine-Dershimer (1978, 1979), nous demanderons aux enseignantes de regrouper les élèves au sein d'ensembles qu'elles estiment hétérogènes, mais dont elles estiment que les élèves qui les composent sont homogènes. La question est ici de vérifier d'une autre manière la pertinence du recours à l'analyse par grappes de scores de jugements pour identifier les catégories d'élèves contenues dans le bagage cognitif des enseignants. Ici, nous comparerons les appartenances aux catégories de la mise en grappes de jugements à celles créées par regroupement d'étiquettes à la façon de Morine-Dershimer (1978).

Après la description de notre échantillon, de notre méthode, de nos traitements statistiques et de nos résultats, nous clôturerons notre contribution par une discussion sur la structure interne des représentations des enseignants et sur les méthodes pour l'analyser.

Méthode

Les participants

Neuf enseignantes ayant entre 6 et 31 années d'expérience (moyenne de 16,5 ans) ont participé à notre recherche composée de trois études (voir Tableau 1). Quatre enseignantes interviennent dans les niveaux inférieurs

de la scolarité élémentaire (maternelle et primaire inférieur), tandis que cinq d'entre elles ont la charge de classes se situant dans les niveaux moyen et supérieur de la scolarité primaire. La taille des différentes classes varie entre 17 et 23 élèves (moyenne de 20 élèves).

Tableau 1
Description du public ayant participé à notre recherche

Enseignantes	Années d'expérience	Degré scolaire actuel	Nombre d'élèves
E1	15	1P	19
E2	10	8P	21
E3	20	3P	23
E4	6	4P	21
E5	10	2P	18
E6	17	1P-2P	17
E7	19	6P	18
E8	31	1P	22
E9	21	7P	22

Nous avons calculé l'effet de l'expérience et du niveau d'enseignement sur base d'une catégorisation des variables d'état (moins et égal à 10 ans contre plus de 10 ans d'expérience ; degré inférieur à 3P contre supérieur et égal à 3P) sur les deux mises en grappes présentées dans la suite (fréquences relatives des grappes). Les tests du Khi2 sont significatifs, mais le calcul du V de Cramer montre que l'effet est très modéré, comme l'indique le tableau 2. D'ailleurs, les analyses des statistiques descriptives ne permettent pas d'identifier de logique sous-jacente à ces différences statistiquement significatives, mais d'amplitude peu prononcée ($V_c < 0,50$).

Tableau 2
Effet de l'expérience et du niveau d'enseignement sur les catégorisations

Effet	Expérience		Niveau d'enseignement	
	χ^2	V_c	χ^2	V_c
Grappes de jugements (ddl=2)	10,5*	0,24*	6,9*	0,19*
Grappes de regroupement (ddl=3)	28,0*	0,39*	24,1*	0,37*

Note. * significatif au seuil 0,05

Les tâches et la procédure

La recherche présentée dans cet article comporte trois phases appelées études dans la suite du compte rendu. Toutes les prises de données ont été faites pour chaque enseignante dans un délai d'un mois. Une fois les données rassemblées (environ deux semaines) pour toutes pour la première étude, nous avons procédé à l'analyse par grappes. Les études 2 et 3 ont par la suite été réalisées simultanément[1] pour les enseignantes à l'intérieur des deux semaines suivantes. À noter que les élèves ont été présentés dans l'ordre alphabétique pour toutes les études.

Étude 1 : jugements des élèves par les enseignantes et analyse par grappes

Comme première tâche, nous avons demandé aux enseignantes de remplir un questionnaire dans lequel elles devaient juger tous leurs élèves selon 23 caractéristiques (voir Tableau 3) à l'aide d'une échelle de type Likert à 7 modalités. Les caractéristiques sont extraites du questionnaire original de Hofer duquel nous avons écarté certaines variables (7 variables sur 23 ; p. ex., l'apparence physique ou le statut socioéconomique familial) pour leur préférer des variables liées aux situations d'enseignement-apprentissage (p. ex., besoin en accompagnement et suivi ou recadrage disciplinaire). En effet, ces variables sont plus pertinentes en ce qu'elles prédisent mieux les interactions et en ce que les enseignants ne rechignent pas à y répondre (Wanlin, 2007a, 2009 ; Wanlin & Crahay, 2011). Une fois ces jugements rassemblés pour toutes les enseignantes de notre échantillon, nous avons procédé, de la même manière que Hofer (1981), à une analyse par grappes pour obtenir différents profils d'élèves définissables par leurs scores moyens aux variables du questionnaire.

Étude 2 : classement des élèves dans les grappes par les enseignantes

Les définitions et étiquettes de grappes trouvées à l'issue de l'étude 1 ont été présentées aux enseignantes lors d'un second entretien. Les enseignantes ont obtenu une feuille avec les étiquettes et les définitions de chaque grappe. (Ces éléments sont reproduits intégralement dans la sous-section Étude 1 : application de la méthode de Hofer à nos données présentée ci-dessous, où les grappes sont décrites grâce à l'analyse de variance.) Les définitions ont été rédigées par les chercheurs sur la base des moyennes des grappes aux scores de variables, puis ces définitions ont

été résumées par les chercheurs par des étiquettes courtes. Cette feuille a été lue par les enseignantes, qui devaient alors classer chacun de leurs élèves à l'intérieur de l'une des grappes, puis estimer dans quelle mesure cet élève est un bon représentant de la grappe choisie. Pour ce faire, nous leur avons demandé d'estimer la proximité de chaque élève à sa grappe d'appartenance en lui attribuant un score variant entre 0 (représentant parfait) et 10 (pas du tout représentatif du profil). Chaque enseignante estimait ensuite, selon la même échelle, dans quelle mesure chaque élève est proche ou éloigné des grappes auxquelles il n'avait pas été associé. L'estimation de la proximité des élèves aux grappes permet notamment d'examiner si les données correspondent à la théorie des grappes isolées organisées de manière centrale-périphérique. En effet, nous conjecturons que, si les enseignantes estiment que les élèves sont plus proches des définitions des grappes auxquelles elles les attribuent, et que les distances aux grappes auxquelles ils ne sont pas attribués sont plus élevées, alors l'hypothèse de l'organisation centrale-périphérique de grappes isolées pourra être conservée.

Étude 3 : regroupements libres des élèves

Lors d'un troisième entretien (qui s'est déroulé après avoir rempli le questionnaire et le même jour que l'entretien d'attribution de grappes), les enseignantes devaient classer leurs élèves en catégories, à l'aide de petits carrés de papier sur lesquels étaient inscrits les noms des élèves. Nous avons emprunté cette tâche à Morine-Dershimer (1978, 1979). Les enseignantes pouvaient composer autant de groupes qu'elles le voulaient.

Nous avons invité les enseignantes à utiliser autant de critères qu'elles le désiraient, puis à décrire chacun des groupes lors de la seconde étape. Nous leur avons également demandé d'estimer le poids de chacun des critères de regroupement, et ce, pour chacun des groupes, à l'aide d'une échelle allant de 0 (peu de poids) à 10 (beaucoup de poids). Nous nous sommes inspirés de Hörstermann et ses collaborateurs (2010), qui ont fait une analyse de contenu réalisée par une analyse par grappes sur la fréquence des adjectifs utilisés par de futurs enseignants pour décrire des types d'élèves. Au lieu de recourir aux fréquences des adjectifs, nous avons utilisé les poids qui leur étaient attribués afin de correspondre à l'hypothèse de hiérarchisation des caractéristiques descriptives émise dans la théorie de la structuration des connaissances (Reed, 2011 ; Sternberg, 2007). Nous avons

additionné ces poids au sein de cinq champs sémantiques (comportement, attention, personnalité, compréhension et application)[2] dichotomisés en valeurs positive et négative. La différence des valeurs positive et négative aboutissait à cinq scores de critères qui ont servi de base pour l'analyse par grappes relative à cette étude. Cela nous a permis d'effectuer, comme Hörstermann et ses collaborateurs (2010), notre analyse de contenu par une mise en grappes informatique.

Pour finir, comme lors de l'étude 2, les enseignantes étaient ensuite invitées à estimer la proximité de chaque élève au groupe qui lui a été attribué sur un score variant entre 0 (représentant parfait) et 10 (pas du tout représentatif du profil), puis à estimer la proximité aux groupes auxquels il n'avait pas été associé.

Les traitements statistiques

Pour les analyses statistiques, nous avons eu recours à des analyses par grappes, comme les auteurs précédents. L'analyse par grappes est une technique statistique qui permet de classer, en fonction de variables données, un ensemble d'objets en sous-ensembles hétérogènes contenant des objets homogènes (Everitt et al., 2011). Il est possible de distinguer les procédures agglomératives et de division. Les premières partent de chaque individu pour les agglutiner un à un, selon leurs ressemblances, jusqu'à l'obtention d'un ensemble comportant tous les individus (la méthode de Ward, dans notre cas, afin d'utiliser le même algorithme principal que les recherches précédentes). Les secondes partent d'un ensemble contenant tous les individus pour séparer progressivement les éléments dissemblables, jusqu'à l'obtention de singletons (analyse en nuées dynamiques, dans notre cas). Pour notre mise en grappes, nous avons suivi les recommandations de Clatworthy, Buick, Hankins, Weinman et Horne (2005). Selon eux, lorsqu'un auteur rapporte une analyse par grappes, il doit :

1. *Mentionner le programme statistique utilisé.* Nous avons utilisé le logiciel SPSS 22 ;
2. *Indiquer la mesure de similarité utilisée.* Dans le cadre de notre recherche, il s'agit du carré de la distance euclidienne (pour les jugements, sur les scores *z* des 23 variables, et pour les descriptions issues des regroupements libres, sur la standardisation des cinq scores obtenus par la soustraction des valeurs positive et négative des champs sémantiques) ;

3. *Rapporter la procédure de détermination des groupes.* Ici, une analyse du dendrogramme et de la chaîne d'agrégation selon laquelle une solution à trois grappes doit être retenue pour la mise en grappes basée sur les jugements (voir Figure 1) et en quatre grappes pour celle basée sur les regroupements libres ;
4. *Confronter l'algorithme agglomératif à un algorithme de division.* Pour la mise en grappes des jugements, nous obtenons un recouvrement de 85,6 %, avec une analyse en nuées dynamiques (nombre de grappes fixé à trois)[3] ;
5. *Vérifier la validité de la mise en grappes et le choix du nombre de groupes.* Ce critère inclut la vérification de la stabilité du modèle et de sa valeur pour le domaine :

a) *Stabilité* : Pour vérifier la stabilité, deux méthodes peuvent être utilisées. Procéder à une nouvelle analyse par grappes en utilisant un autre algorithme hiérarchique agglomératif. C'est la procédure utilisée par Hofer. Nous n'utilisons pas cette procédure, car des principes de regroupement différents aboutissent inévitablement à des grappes différentes (voir Clatworthy et al., 2005), entraînant une sous-évaluation des taux de recouvrement. Nous lui préférons la méthode du partitionnement aléatoire des données en deux parties pour y reproduire la méthode de Ward, à l'origine de la première mise en grappes. Les recouvrements entre les partitions et la mise en grappes totale sont de 91,2 % et 88,9 % ; la comparaison des définitions de partitions obtenues par la comparaison de leurs moyennes aboutit à un recouvrement parfait entre les deux partitions ;

b) *Valeur pour le domaine* : Pour vérifier si la mise en grappes s'explique par une autre ou d'autres variables non incluses dans le processus de regroupement, la suite de notre article apportera un éclairage à ce sujet (voir les recouvrements entre les mises en grappes et l'attribution des enseignantes).

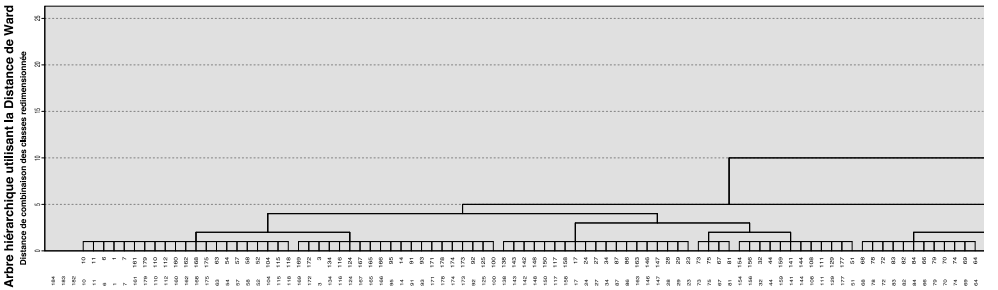

Figure 1. ***Dendrogramme de classification des élèves par la méthode de Ward***

Résultats

Étude 1 : application de la méthode de Hofer à nos données

Nous aboutissons donc à une répartition des élèves en trois grappes (voir Figure 1) qui survit bien aux critères de qualité de Clatworthy et ses collaborateurs (2005). Ces grappes peuvent être définies sur la base de leurs scores moyens aux différentes variables obtenues par des analyses de variance, et par des tests *post-hoc* (selon la méthode de Scheffé) comparant les moyennes une à une en vue d'éviter d'accepter ou de rejeter des différences significatives alors qu'elles existent ou non (Howell, 1998). Ces comparaisons sont disponibles dans le tableau 3.

Les 111 élèves de la grappe J1 ont des scores d'intérêt et de capacités d'apprentissage meilleurs que ceux des autres grappes. Leur niveau scolaire est jugé très bon (en mathématiques et en lecture, mais aussi de manière générale). Les enseignantes pensent qu'ils se comportent très bien en classe, sont plus appliqués, polis et obéissants, et qu'ils ont moins besoin de recadrage disciplinaire. Ces élèves sont perçus comme étant ouverts, sociables et équilibrés. Cette grappe de *(très) bons élèves* s'apparente aux deux premières grappes identifiées par Hofer (1981)[4].

La grappe J2 contient 40 élèves que les enseignantes jugent peu intéressés et aux capacités d'apprentissage peu élevées. Les enseignantes estiment que leur niveau scolaire est bas (pour les mathématiques, la lecture et de manière générale) et qu'ils ont un besoin élevé de guidage. Leur comportement en classe est perçu comme étant moyen avec du recadrage disciplinaire proche de ceux de la grappe J1. Cette grappe, perçue plutôt néga-

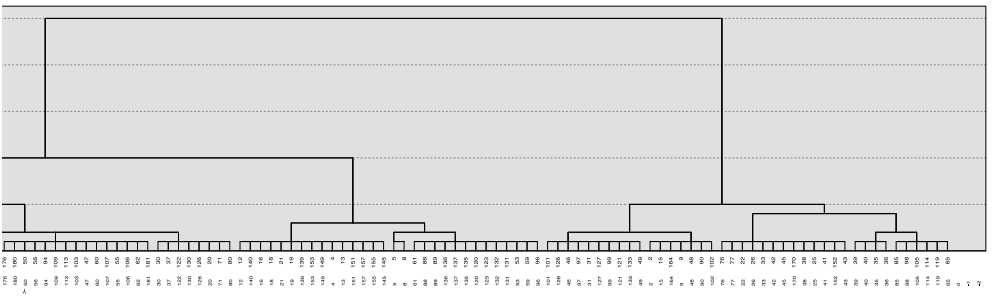

Tableau 3

Comparaison des moyennes obtenues par les différentes grappes d'élèves aux variables jugées par les enseignantes (ANOVA avec tests post-hoc)

Variables (n° d'item)	J1 (111 élèves)		J2 (40 élèves)		J3 (30 élèves)		Total (181 élèves)		ANOVA		
	M	ET	M	ET	M	ET	M	ET	F	sig.	Scheffé
Intérêt (1)	6,2	0,9	4,2	1,1	5,2	1,1	5,6	1,3	64,772	0,000	2 < 3 < 1
Application (2)	6,3	0,8	4,8	1,4	4,4	0,9	5,6	1,3	63,262	0,000	2 = 3 < 1
Attention (3)	6,3	1,0	3,9	1,4	4,5	0,9	5,5	1,5	81,810	0,000	2 = 3 < 1
Comportement (4)	6,3	1,2	4,8	2,3	3,7	1,3	5,5	1,8	40,527	0,000	3 < 2 < 1
Obéissance (5)	6,6	0,6	5,7	1,7	3,7	1,5	5,9	1,5	81,082	0,000	3 < 2 < 1
Niveau général (6)	6,3	0,9	3,5	1,3	5,5	1,0	5,5	1,5	109,883	0,000	2 < 3 < 1
Niveau de lecture (7)	6,0	1,1	3,4	1,5	5,5	1,0	5,4	1,6	70,600	0,000	2 < 3 = 1
Niveau en mathématiques (8)	6,2	0,9	4,0	1,6	5,5	1,0	5,6	1,4	53,679	0,000	2 < 3 < 1
Capacité d'apprentissage (9)	6,5	0,8	4,3	1,3	5,8	1,0	5,9	1,3	79,424	0,000	2 < 3 < 1
Recadrage disciplinaire (10)	2,5	2,2	3,2	2,4	4,6	1,5	3,0	2,3	11,312	0,000	1 = 2 < 3
Guidage explicatif (11)	1,7	1,0	4,8	1,7	2,7	0,9	2,6	1,7	107,322	0,000	1 < 3 < 2
Concentration (12)	6,3	0,9	3,8	1,4	4,5	1,0	5,5	1,5	101,249	0,000	2 < 3 < 1
Accompagnement et suivi (13)	1,6	0,9	4,9	1,8	2,2	0,7	2,4	1,8	123,870	0,000	1 < 3 < 2
Politesse (14)	6,8	0,7	6,1	1,4	5,6	1,3	6,4	1,1	18,506	0,000	3 = 2 < 1
Équilibre (15)	6,6	0,7	4,6	1,5	5,7	1,3	6,0	1,3	51,986	0,000	2 < 3 < 1
Sympathie (16)	6,5	0,6	5,8	0,9	5,8	1,3	6,2	0,9	16,319	0,000	2 = 3 < 1
Renfermement (17)	5,9	1,3	4,1	1,5	6,4	0,9	5,6	1,5	31,783	0,000	2 < 1 = 3
Confiance en soi (18)	5,7	1,3	2,9	1,4	6,4	0,8	5,2	1,8	85,735	0,000	2 < 1 < 3
Sociabilité (19)	6,4	0,8	4,9	1,6	6,4	1,1	6,1	1,2	26,809	0,000	2 < 1 = 3
Besoin de se faire valoir (20)	5,2	1,6	4,6	2,0	4,1	1,9	4,9	1,8	5,124	0,007	(3 = 2) < (2 = 1)
Leadership (21)	4,6	1,7	3,6	1,4	4,8	1,6	4,4	1,7	6,345	0,002	2 < 1 = 3
Valorisations (22)	3,8	1,7	5,8	1,4	4,2	1,5	4,3	1,8	23,684	0,000	1 = 3 < 2
Feedbacks (23)	3,3	1,8	4,4	2,0	3,9	1,5	3,7	1,9	5,896	0,003	(1 = 3) < (3 = 2)

Note. M = moyenne; ET = écart-type.

tivement aux items psychologiques et sociaux, renferme des *élèves en difficulté* que nous pourrions comparer à la grappe 3 identifiée par Hofer (1981), voire éventuellement à sa grappe 5.

La grappe J3 comporte 30 élèves qui, bien que leur niveau en lecture soit proche de celui des élèves de la grappe J1, ont des scores moyens aux variables scolaires (intérêt, niveau en mathématiques et de manière générale). Les enseignantes pensent qu'ils ont parfois besoin de plus de guidage que les élèves de la grappe J1, mais moins que ceux de la grappe J2. Leur comportement en classe est jugé peu propice à l'apprentissage appliqué et ils obtiennent, selon les enseignantes, davantage de recadrage disciplinaire que les deux autres grappes. Leurs scores aux éléments psychologiques et sociaux sont proches de ceux de la grappe J1. Cette grappe de *perturbateurs moyens* peut être rapprochée de la grappe 4 identifiée par Hofer (1981).

La vérification de la pertinence de la méthode de mise en grappes de scores de jugements

Cette section de l'article porte successivement sur les données des études 2 et 3.

Données de l'étude 2 : classement des élèves dans les grappes par les enseignantes

Dans cette section, nous analysons si les enseignantes classent leurs élèves dans la même grappe que l'algorithme informatique. Pour rappel, nous avons, pour ce faire, proposé les définitions de grappes susmentionnées (avec les étiquettes) et demandé aux enseignantes d'attribuer une grappe à chacun de leurs élèves. Cette procédure, illustrée au tableau 4, aboutit à un classement identique entre l'attribution de l'enseignante et la mise en grappes informatique pour 64,6 % des élèves, c'est-à-dire 117 d'entre eux ($\chi^2_{(4)} = 45{,}39$; $p < 0{,}001$; $V_c = 0{,}501$; $p < 0{,}001$). Pour les autres 64 élèves (35,4 %), les classifications peuvent être de deux types : des classifications moins favorables (p. ex., 14 élèves de la grappe J1 sont classés dans la grappe J2, tandis que 11 élèves le sont dans la grappe J3) et des classifications plus favorables (p. ex., 12 élèves de la grappe J3 sont étiquetés comme de bons élèves).

Le tableau 5 présente les taux de recouvrement pour chaque enseignante entre son attribution de grappes et le classement obtenu de manière informatique. La fourchette de recouvrement oscille entre 50 % et 91 %. Les enseignantes ne sont donc pas égales en matière de conformité de leur attribution d'étiquettes avec les classifications obtenues informatiquement.

Tableau 4
Comparaison des attributions de grappes par l'algorithme statistique et par les enseignantes

Grappe attribuée informatiquement	Étiquette grappe attribuée par l'enseignante J1	J2	J3	Total
J1	86	14	11	111
J2	11	18	11	40
J3	12	5	13	30
Total	109	37	35	181

Tableau 5
Concordance entre les attributions de grappes par les enseignantes et l'attribution obtenue de manière informatique (pourcentage par enseignante)

N° d'enseignante	Taux de recouvrement brut	$\chi^2_{(4)}$	V_c	Remarque
1	0,68	12,06**	0,56**	
2	0,52	4,23	0,32	
3	0,70	18,30***	0,63***	
4	0,76	11,20***a	0,73***	c
5	0,50	8,44*	0,48*	
6	0,59	0,37a	0,15	c
7	0,50	5,72*a	0,56*	d
8	0,59	11,48**	0,51**	
9	0,91	9,90***b	0,67***	e

Note. a = $\chi^2_{(2)}$; b = $\chi^2_{(1)}$; c = aucun élève classé en C3 par l'ordinateur; d = aucun élève classé en C3 par l'enseignante; e = aucun élève classé en C3 ni par l'ordinateur ni par l'enseignante.

Il est difficile d'apprécier si le taux de recouvrement global d'environ 65 % (variant entre 50 % et 91 % selon l'enseignante) est satisfaisant ou non et de conclure que les grappes générées de manière informatique par la méthode de Hofer représentent fidèlement les catégories que les enseignantes possèdent sur leurs élèves. Néanmoins, il ne faut pas non plus rejeter cette éventualité puisque le test du χ^2 et le V de Cramer indiquent qu'il y a une logique dans les associations et que leur lien est relativement robuste.

Données de l'étude 3 : analyse des regroupements libres des élèves et de leur recouvrement avec les grappes obtenues dans l'étude 1

Le système de scores que nous avons obtenu a été soumis à une analyse par grappes (comme l'ont fait Hörstermann et al., 2010) pour obtenir un nombre de groupes identiques puisque les enseignantes ont donné des systèmes de deux à six ensembles. La mise en grappes que nous avons utilisée pour cette étude respecte aussi les critères de Clatworthy et ses collaborateurs (2005). Les analyses de variance et les tests *post-hoc* fournis au tableau 6 permettent de décrire une catégorisation des élèves en quatre ensembles.

Tableau 6

Comparaison des moyennes obtenues par les différentes grappes d'élèves aux scores de champs sémantiques (ANOVA avec tests post-hoc)

	G1	G2	G3	G4	$F_{3/34}$	sig.	Scheffé 0,05 /0,10
Comportement	-0,08	-21,30	-7,50	-2,22	13,11	0,000	2 < (3=4=1)
	(9,06)	(10,40)	(7,64)	(5,38)			
Attention	-2,23	8,50	0,33	2,33	3,56	0,024	/
	(4,62)	(10,59)	(5,47)	(9,54)			
Personnalité	2,15	-8,00	9,67	-18,11	17,91	0,000	(2=) 4 < 3 (=1 = 2)
	(7,50)	(5,25)	(13,32)	(7,46)			
Compréhension	5,46	-7,40	4,17	-2,33	7,48	0,001	(4=) 2 < (3=1) (=4)
	(0,09)	(6,04)	(6,21)	(4,12)			
Application	0,08	2,20	14,17	7,22	14,18	0,000	(2=) 1 < 4 (=2) < 3
	(4,40)	(4,92)	(5,46)	(4,29)			

Note. M = moyenne; ET = écart-type.

La grappe G1 comporte des groupes d'élèves décrits favorablement, hormis concernant l'application. Nous pourrions appeler cette grappe les *élèves décrits positivement mais peu appliqués*. La grappe G2 rassemble des groupes d'élèves perturbateurs à personnalité non conforme aux attentes scolaires – ils sont provocateurs ou excités – et présentant des résultats scolaires peu élevés. Il s'agit des *agités et peu performants*. La grappe G3 contient des groupes d'élèves parmi les plus appliqués présentant une personnalité conforme aux attentes scolaires et des résultats scolaires positifs, bien qu'ils soient quelque peu agités de temps à autre. L'étiquette *appliqués et positivement perçus* lui convient. La grappe G4 renferme des groupes d'élèves à la personnalité la moins conforme aux attentes sco-

laires – ils sont effacés et participent peu –, mais qui présentent les descriptions les plus positives concernant l'implication dans les tâches scolaires. L'appellation *discrets et impliqués* convient à cette grappe.

L'analyse suivante consiste à vérifier si les regroupements effectués par les enseignantes et les grappes générées informatiquement concordent entre eux. D'un point de vue purement théorique, il apparaît, de l'analyse des tableaux 3 et 6, que des rapprochements peuvent être faits entre les deux méthodes de classification. Nous suggérons le résultat de cette analyse théorique dans le tableau 7. D'un point de vue parfaitement conceptuel, la grappe J1 peut être rapprochée des ensembles des *élèves décrits positivement mais peu appliqués* (G1) et des *appliqués et positivement perçus* (G3) (concordance acceptée). De plus, une association timide peut être vue avec l'ensemble des *discrets et impliqués* (G4) (concordance acceptable), mais une connexion avec l'ensemble des élèves *agités et peu performants* (G2) est exclue (concordance refusée). Pour les grappes J2 et J3, les concordances acceptées sont celles avec les ensembles des élèves *agités et peu performants* (G3) et *discrets et impliqués* (G4). Les concordances acceptables incluent l'ensemble des *bons élèves* (G2). Une association avec les *bons élèves peu appliqués* (G1) est conceptuellement exclue.

Une relation entre les regroupements et la mise en grappes est effective ($\chi^2_{(6)} = 23{,}20$; $p < 0{,}001$), mais peu importante ($V_c = 0{,}25$; $p < 0{,}001$). Les taux de recouvrement oscillent entre 49 % pour les recouvrements acceptés et 63,4 %, en incluant également les recouvrements acceptables. Les recouvrements ne sont cependant pas identiques pour chaque grappe. Les classifications des élèves perçus positivement concordent davantage (entre 57 % et 74 %) que pour les élèves perçus moins favorablement (environ 60 % pour les élèves en difficulté et moins de 30 % pour les perturbateurs).

Tableau 7

Comparaison entre la mise en grappes obtenue par les scores de jugements (étude 1) et celle obtenue par la tâche de regroupement (étude 3)

	G1 élèves décrits positivement mais peu appliqués			**G2 élèves agités et peu performants**			**G3 élèves appliqués et positivement perçus**			**G4 élèves discrets et impliqués**			**Classements concordants**
	C+	R+	A-	C-	R-	Ao-	Co+	R(o)+	A+	Co	Ro-	Ao+	(%)
J1 : (Très) bons élèves		Accepté			Refusé			Accepté			Acceptable		
C+	■									▪			Classements acceptés
R+		■						■					56,76
A+									■			▪	Acceptables inclus
Nombre d'élèves		40			29			23			19		73,87
J2 : En difficulté		Refusé			Accepté			Acceptable			Accepté		
Co							▪			■			Classements acceptés
R-					■						■		57,50
A-						■							Acceptables inclus
Nombre d'élèves		16			13			1			10		60,00
J3 : Perturbateurs moyens		Refusé			Accepté			Acceptable			Accepté		
C-				■			▪						Classements acceptés
Ro								▪			■		10,00
Ao						■						■	Acceptables inclus
Nombre d'élèves		21			3			6			0		30,00

Note. C = comportement; R = résultats scolaires, performance et compréhension; A = attention/application; + = positif; o = moyen; - = négatif;
■ = recouvrement analogue; ▪ recouvrement partiel.

La vérification de l'hypothèse de l'organisation centrale-périphérique des grappes

La perception de la typicalité des élèves aux grappes

a. Données de l'étude 2 : grappes de jugements

Nous avons demandé aux enseignantes d'estimer dans quelle mesure chaque élève était un bon représentant de la grappe à laquelle il est associé et dans quelle mesure il se distancie des autres grappes. Les données que nous avons rassemblées dans ce contexte peuvent être résumées à l'aide du tableau 8.

Tableau 8
Moyennes des distances aux grappes de jugements (ANOVA et tests post-hoc)

Attribution d'étiquette par l'enseignante	J1		J2		J3				
	M	ET	M	ET	M	ET	$F_{2/178}$	sig.	Scheffé
Distance à la grappe d'appartenance	1,5	1,5	2,6	2,2	2,1	2,0	6,39	0,002	(1 = 3) < (3 = 2)
Distance aux autres grappes	8,2	1,6	5,8	1,6	4,7	1,7	76,68	0,000	3 < 2 < 1
F	$_{1/216}$ = 1012,5		$_{1/72}$ = 106,20		$_{1/68}$ = 69,80				
sig.	0,000		0,000		0,000				

Note. M = moyenne; ET = écart-type.

Pour les enseignantes, les élèves catégorisés dans la grappe J1 sont plus proches de leur groupe d'appartenance et plus éloignés des autres grappes. Pour les élèves des grappes J2 et J3, les distances moyennes à leur grappe d'appartenance sont généralement plus importantes, tandis que leurs distances aux autres grappes sont plus petites que celles estimées pour les élèves de la grappe J1.

b. Données de l'étude 3 : grappes de regroupement

L'estimation des distances de chaque élève aux groupes constituait aussi une part de cette étude. Le tableau 9 montre que les distances aux grappes d'appartenance sont plus resserrées que les distances aux autres grappes, même si elles sont plus comprimées pour la grappe G3 que pour la G1. Les élèves classés dans la grappe G1 sont donc plus représentatifs de leur groupe que ne le sont les élèves classés dans G3. Les dissimilarités aux autres grappes sont plus étendues pour la G4 et plus étroites pour la G1.

Tableau 9
Moyennes des distances aux grappes de jugements (ANOVA et tests post-hoc)

	G1	G2	G3	G4	F	dl	sig.	Scheffé
Distance au groupe d'appartenance	4,84 (1,59)	4,07 (1,64	3,23 (1,33)	3,97 (1,78)	7,13	3/177	0,000	(4=2=) 3 < 1 (=2=4)
Distance aux autres groupes	6,53 (2,49)	7,24 (1,67)	7,24 (1,70)	8,52 (1,88)	6,09	3/140	0,001	(2=3=) 1 < 4 (=2=3)
F	19,82	82,93	131,07	68,33				
dl	1/115	1/88	1/58	1/56				
sig.	0,000	0,000	0,000	0,000				

Note. M = moyenne; ET = écart-type.

Études 2 et 3 : comparaison des degrés de typicité des élèves

La comparaison entre les distances aux grappes de jugements et les distances aux grappes de regroupement, illustrée dans la figure 2, montre que la mise en grappes informatique (en noir dans la figure) tend à maximiser la ressemblance des élèves appartenant à une même grappe et à maximiser les distances entre grappes. Les distances issues des regroupements proposés par les enseignantes (en gris dans la figure) aboutissent à des ensembles plus amples en matière de ressemblance, mais

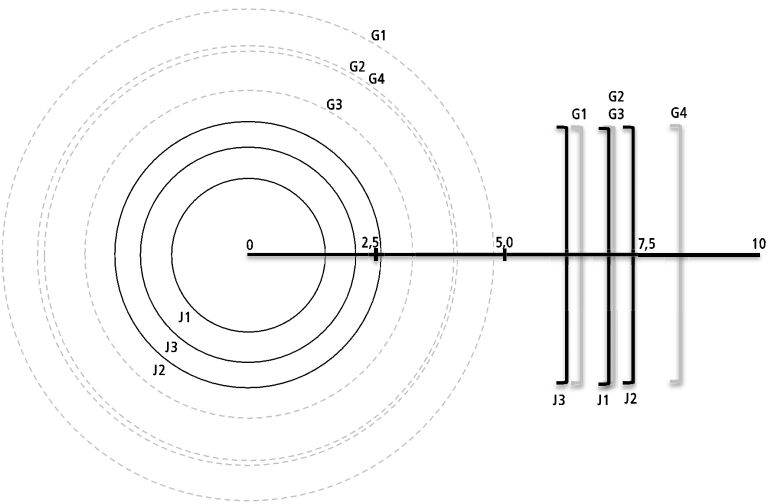

Figure 2. ***Comparaison, par superposition, des distances aux grappes de jugements ou de regroupement. En cercle = les proximités à la grappe attribuée ; en segment = les distances aux autres grappes ; en traits pleins et noirs = les grappes de jugements ; en traits pointillés et gris = les grappes de regroupement.***

aussi à des ensembles perçus comme étant moins distants les uns des autres. Exprimé autrement, l'algorithme statistique tend à amplifier la typicité des élèves dans leur grappe. Concernant l'organisation centrale-périphérique, nos résultats ne semblent pas nous permettre, au moins pour les élèves présentant les caractéristiques les plus typiques des prototypes de leurs grappes de jugements ou de regroupement, de la mettre en cause.

Discussion et conclusion

Nos analyses ont permis de mettre en évidence trois ou quatre profils d'élèves selon la méthode utilisée. Les trois profils issus de la mise en grappes des jugements peuvent être rapprochés des cinq ensembles mis en évidence par Hofer (1981), qui a utilisé cette même méthode. Afin de vérifier la présence effective de ces profils dans l'esprit des enseignantes, nous leur avons demandé de classer leurs élèves à l'intérieur de ces trois ensembles. Le taux de recouvrement est de 65 %. Une autre tâche, partie d'un regroupement libre, aboutit à quatre ensembles. Les associations des grappes de jugements avec ces grappes de regroupement oscillent entre 50 % et 65 %. La méthode de Hofer ne représente donc pas de manière totalement fidèle les catégories que nos enseignantes possèdent sur les élèves. Néanmoins, la méthode n'est pas totalement dénuée de sens. Cette observation rejoint les constats de Hofer et Köpke (1987).

L'analyse de la proximité des élèves aux différentes grappes estimée par les enseignantes permet des observations intéressantes. Pour rappel, Hofer conjecturait la présence de catégories d'élèves qui adoptent chacune la structure d'un noyau central autour duquel gravitent les différents élèves ; ces catégories sont également hermétiques entre elles. Cette hypothèse est devenue tellement classique qu'elle est tenue pour vraie par la plupart des chercheurs. Or, nous n'avons pas pu recenser d'études publiées et accessibles qui vérifient cette position communément prise pour correcte. Notre examen met en lumière que cette structure est possible. Ainsi, les distances aux grappes attribuées par les enseignantes sont plus petites que celles qui renvoient aux autres grappes, c'est-à-dire que, lorsqu'une enseignante attribue une étiquette de grappe à un élève, elle l'estime plus proche de cette description que des descriptions correspondant aux autres grappes identifiées à l'aide de l'algorithme statistique. L'identification de trois grappes distantes les unes des autres peut signifier que des catégories distinctes d'élèves coexistent dans le bagage cognitif de nos enseignantes,

et la proximité plus petite pour les élèves associés à une même grappe peut signifier que les grappes ont des noyaux centraux autour desquels gravitent les différents élèves. Nous pensons cependant que nos constats embrassent la conception de la structure interne des croyances que proposait Green (1971), plutôt que celle proposée par Hofer (1986), qui est probablement le fruit d'un artifice statistique, comme l'a suggéré Oldenbürger (1986). Rappelons que la différence entre les deux conceptions résidait dans la perméabilité des frontières entre grappes. En effet, l'observation des données indique que certains élèves obtiennent des distances limites (plus éloignées de leur grappe et moins éloignées des autres grappes), ce qui pourrait mettre à mal l'hypothèse de la perméabilité des frontières des catégories d'élèves. Par ailleurs, l'attribution de distances différentes aux élèves classés librement permet d'étayer cette observation. En effet, il n'était pas rare qu'un élève obtienne une distance de 6 dans son groupe et de 5 dans les autres groupes.

Même si le système quantitatif de Hofer n'est pas parfait pour identifier les catégories d'élèves, il est relativement intéressant s'il est combiné avec une approche qualitative d'attribution d'étiquettes et d'estimation de proximité aux grappes. Dans notre cas, nous pouvons envisager l'utilisation d'une échelle de jugements comme un point de départ et effectuer la catégorisation informatique par une étape de consultation des enseignants (attribution et estimation de proximité). Toutefois, il faut rester attentif à deux éléments dans l'interprétation des données. Premièrement, d'après nos analyses, l'algorithme de mise en grappes à partir des jugements émis par les enseignantes amplifie la ressemblance entre les élèves, alors que les enseignantes estiment qu'ils se distinguent davantage les uns des autres lors des associations libres. Deuxièmement, il n'est pas exclu que le constat selon lequel les enseignantes attribuent des distances plus élevées aux grappes auxquelles elles n'associent pas leurs élèves et des distances plus petites pour celles auxquelles elles les associent ne soit pas le fruit d'une sorte d'effet de désirabilité sociale. Ainsi, vu que les enseignantes devaient classer chaque élève dans une grappe, puis estimer sa proximité à toutes les grappes, il est possible qu'elles aient tenté de rester consistantes avec leur attribution ou non-attribution. Une manière de procéder serait de demander un degré de typicité sans précéder cet exercice d'une attribùution de grappes. Par ailleurs, pour limiter le nombre de traitements statistiques, il aurait été intéressant de compléter les regroupements libres par une seconde étape dans laquelle une restriction quant au nombre d'ensembles serait

imposée. En effet, les enseignantes ont proposé un nombre différent de groupes d'élèves, ce qui a vraisemblablement eu une incidence sur l'ampleur des recouvrements que nous avons identifiés.

En définitive, notre recherche indique que l'approche de Hofer prise seule s'avère insuffisante pour l'analyse de la structure des représentations des enseignants en matière de catégories d'élèves ; il faut la compléter notamment par la consultation des enseignants qui ont jugé les élèves. Nos résultats montrent qu'il se pourrait que les représentations des enseignants concernant leurs élèves soient organisées, pour certains élèves, en grappes dont l'organisation interne s'apparenterait à un noyau central autour duquel gravitent des éléments périphériques. Cependant, ils montrent aussi que d'autres élèves sont moins typiques de leurs grappes. Il conviendrait de vérifier leur positionnement dans les catégories d'élèves et donc de tester l'éventualité d'une gravitation à l'intersection entre plusieurs catégories. Évidemment, d'autres recherches devraient être entreprises afin de vérifier la pertinence de cette hypothèse.

Les limites de notre étude sont vraisemblablement nombreuses : la taille réduite de l'échantillon ou son hétérogénéité (bien que l'effet du niveau d'enseignement soit modeste). Pour remédier à ces limites, notre équipe conduit actuellement des analyses auprès de 40 enseignants du primaire (n=20) et du secondaire (n=20). Une autre limite réside dans la non-prise en compte des conséquences des catégorisations sur l'interaction en classe et, surtout, du degré de typicité des élèves. Il faudrait aussi vérifier quels facteurs permettent d'influencer ce lien. Nos analyses actuelles comprennent donc également des observations de terrain ainsi que des mesures des connaissances psychopédagogiques des enseignants. D'autres études pourraient analyser si les enseignants se font une représentation plus précise des élèves plus typiques et une représentation plus floue des élèves moins typiques. À nouveau, il s'agirait de vérifier quels facteurs influencent ces représentations et sur quels éléments ils influent. Ces questions sont importantes, car Hofer (1981) montre une influence des grappes sur la quantité et la qualité des interactions en classe. Les méta-analyses actuelles identifient les mêmes résultats (Nurmi, 2012) : les élèves moins favorablement perçus sont au centre d'interactions moins riches pédagogiquement. Par ailleurs, les recherches sur la justesse des compétences diagnostiques (Praetorius, Karst, Dickhäuser & Lipowsky, 2011 ; Südkamp, Kaiser & Möller, 2012) indiquent qu'elles sont imprécises et inexactes. Ce

constat est d'autant plus vrai pour les élèves présentant les profils les moins typiques, car ils sont systématiquement sous-évalués ou surévalués (Canal, Bonini, Micciolo & Tentori, 2012 ; Soodla & Kikas, en préparation).

Nous espérons donc avoir donné envie à d'autres chercheurs de nous accompagner dans l'analyse de ces hypothèses.

Réception : 14 août 2012

Version finale : 11 décembre 2015

Acceptation : 15 février 2016

NOTES

1. Ces études ont été réalisées le même jour, mais ont été inversées pour la moitié de l'échantillon. La contamination d'une étude sur l'autre est modérée ($V_c < 0{,}30$).
2. Les cinq champs sémantiques retenus sont le fruit d'une analyse de contenu (Bardin, 1977 ; Wanlin, 2007b) de l'ensemble des fiches produites par les enseignantes. Cette analyse de contenu a été menée de concert par trois des auteurs de cet article. Concrètement, les descripteurs utilisés pour définir les groupes ont été classés selon leur champ sémantique. Chaque champ sémantique a été organisé selon la valeur positive ou négative des descripteurs. Par exemple, pour le champ sémantique « comportement », nous pouvions distinguer les adjectifs « sage » (positif) et « perturbateur » (négatif) et des synonymes (calme, obéissant ou docile par rapport à agitateur ou provocateur). Chaque descripteur ayant obtenu un poids de la part des enseignantes, nous avons additionné les poids de chaque descripteur rangés dans les pôles négatif et positif des champs sémantiques pour obtenir 10 sous-scores. Les sous-scores positifs et négatifs de chaque champ sémantique ont été soustraits pour obtenir les cinq scores que nous avons utilisés pour la mise en grappes. La raison de la soustraction réside dans le fait que certains sous-scores obtenaient la valeur de 0 marquant l'absence de descripteur ; cette absence est considérée comme une donnée manquante et entraîne l'ignorance du descriptif dans la mise en grappes. La soustraction permet de conserver tous les descriptifs dans l'analyse, tout en tenant compte de la réalité que l'absence d'un sous-score par exemple positif n'entraînait pas forcément l'absence d'un sous-score négatif ; toutes les fiches ont ainsi pu être conservées lors de l'analyse.
3. Les recouvrements pour la mise en grappes basée sur les descriptions des regroupements libres oscillent entre 63 % et 82 %. Pour ne pas surcharger cet article, nous n'en mentionnons pas davantage. Ils sont néanmoins disponibles auprès de l'auteur de référence.
4. Ce paragraphe et les deux suivants correspondent aux définitions et aux étiquettes de grappes qui ont été soumises aux enseignantes dans le cadre de l'étude 2 (l'enseignante classe ses élèves dans les grappes « statistiques »).

BIBLIOGRAPHIE

Bardin, L. (1977). *L'analyse de contenu*. Paris : PUF.

Canal, L., Bonini, N., Micciolo, R., & Tentori, K. (2012). Consistency in teachers' judgments. *European Journal of Psychology of Education*, *27*, 319-327. doi: 10.1007/s10212-011-0073-6

Clatworthy, J., Buick, D., Hankins, M., Weinman, J., & Horne, R. (2005). The use and reporting of cluster analysis in health psychology: A review. *Britsh Journal of Health Psychology*, *10*, 329358. doi: 10.1348/135910705X25697

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). London: Wiley.

Friedrich, G. (1979). Die Wahrnehmung der Schülerpersönlichkeit. In G. Friedrich & E. E. Krainz (Hgb.), *Lehrereinstellungen, Lehrerverhalten, Schulangst*, Pädagogik der Gegenwart. Wien: Jugend & Volk.

Green, T. (1971). *The activities of teaching*. New York: McGraw-Hill.

Hofer, M. (1969). *Die Schülerpersönlichkeit im Urteil des Lehrers*. Weinheim: Beltz.

Hofer, M. (1981). Schülergruppierungen in Urteil und Verhalten des Lehrers. In M. Hofer (Hgb.), *Informationsverarbeitung und Entscheidungsverhalten von Lehrern* (p. 192-221). München: Urban & Schwarzenberg.

Hofer, M. (1986). *Sozialpsychologie erziherischen Handelns: Wie das Denken und Verhalten von Lehrern organisiert ist*. Göttingen: Verlag für Psychologie, Hogrefe.

Hofer, M., & Köpke, B. (1987). Die Kategorisierung von Schülern durch Lehrer, eine Methodenstudie. *Unterrichtswissenschaft, 15,* 285-294.

Hörstermann, T., Krolak-Schwerdt, S., & Fischbach, A. (2010). Die kognitive Repräsentation von Schülertypen bei angehenden Lehrkräften – Eine typologische Analyse. *Revue suisse des sciences de l'éducation*, *32*(1), 143-158. Repéré à http://rsse.elearninglab.org/wp-content/uploads/2012/12/SZBW_10.1_Varia_Hoestermann.pdf

Howell, D. C. (1998). *Méthodes statistiques en sciences humaines*. Bruxelles : De Boeck.

Medin, D. L. (2004). Concepts and conceptual structure. In D. A. Balota & E. J. Marsh (Eds.), *Cognitive psychology: Key readings* (pp. 482-498). New York: Psychology Press.

Morine-Dershimer, G. (1978). How teachers "see" their pupils. *Educational Research Quarterly*, *3*(4), 43-52.

Morine-Dershimer, G. (1979). *Teachers' conceptions of pupils – An outgrowth of instructional context: The South Bay Study, Part III.* Report No. 59, East Lansing, MI: Institute for Research on Teaching, Michigan State University. Retrieved from http://education.msu.edu/irt/PDFs/ResearchSeries/rs059.pdf

Nurmi, J.-E. (2012). Students' characteristics and teacher-child relationships in instruction: A meta-analysis. *Educational Research Review*, *7*, 177-197. doi: 10.1016/j.edurev.2012.03.001

Oldenbürger, H. A. (1986). Does a tendency to group pupils on attributes exist in teachers' cognitions/judgements? In M. Ben Peretz, R. Bromme, & R. Halkes (Eds.), *Advances of research on teacher thinking* (pp. 186-200). Lisse, Netherlands: Swets & Zeitlinger B.V.

Praetorius, A.-K., Karst, K., Dickhäuser, O., & Lipowsky, F. (2011). Wie gut schätzen Lehrer die Fähigkeitsselbstkonzepte ihrer Schüler ein? Zur diagnostischen Kompetenz von Lehrkräften. *Psychologie in Erziehung und Unterricht*, *58*, 81-91.

Reed, S. K. (2011). *Cognition : théories et applications*. Bruxelles : De Boeck.

Rosch, E. (1973). Natural categories. *Cognitive Psychology*, *4*, 328-350. doi: 10.1016/0010-0285(73)90017-0

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, *7*, 532-547. doi: 10.1016/0010-0285(75)90021-3

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.

Soodla, P., & Kikas, E. (in production). Teachers' judgments of students' reading difficulties and factors related to its accuracy. In A. Toomela (Ed.), *Systematic person-oriented study of child development in early primary school*. Frankfurt am Main: Peter Lang.

Sternberg, R. J. (2007). *Manuel de psychologie cognitive : du laboratoire à la vie quotidienne*. Bruxelles : De Boeck.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(3), 743-762. doi: 10.1037/a0027627

Thelen, H. A. (1967). *Classroom grouping for teachability*. New York: John Wiley.

Wanlin, P. (2007a). Présentation des indices qu'une enseignante de première année primaire avance pour expliquer ses prises de décision lors de la phase interactive de son enseignement de la lecture. *Cahiers des sciences de l'éducation*, *23-24*, 19-38.

Wanlin, P. (2007b). L'analyse de contenu comme méthode d'analyse qualitative d'entretiens : une comparaison entre les traitements manuels et l'utilisation de logiciels. *Recherches qualitatives* (hors série), *3*, 243-272. Repéré à http://www.recherche-qualitative.qc.ca/documents/files/revue/hors_serie/hors_serie_v3/Wanlin2.pdf

Wanlin, P. (2009). High or low performance pupils: Who do steer the instructional rhythm? Paper presented at the EARLI Biennial Conference, Amsterdam: University of Amsterdam.

Wanlin, P., & Crahay, M. (2011). Les enseignants utilisent-ils leurs connaissances ou perceptions des élèves lorsqu'ils donnent cours ? *Dossiers des sciences de l'éducation*, *26*, 51-64.