

Catégorisation des élèves par les enseignants : comparaison de trois méthodes de recherche¹

Philippe Wanlin – philippe.wanlin@unige.ch ou philippe_wanlin@hotmail.com
Université de Genève, Suisse

Lara Laflotte – lara.laflotte@hepl.ch
Haute École Pédagogique du canton de Vaud, Suisse

Pour citer cet article : Wanlin, P., & Laflotte L. (2020). Catégorisation des élèves par les enseignants : comparaison de trois méthodes de recherche. *Évaluer. Journal international de recherche en éducation et formation*, 6(2), 3-22.

Résumé

La littérature postule l'existence de catégories d'élèves que les enseignants utilisent pour préparer et donner leurs leçons. Différentes méthodes sont employées pour les évaluer et trois d'entre elles sont plus particulièrement investiguées ici : le questionnaire de jugements, les descriptifs de types d'élèves et les regroupements libres d'élèves. Nous exposons ces méthodes et proposons une recherche qui les intègre afin d'évaluer les caractéristiques des catégories d'élèves recueillies par chacune d'elles, auprès de quarante enseignants du primaire et du secondaire du canton de Genève, en vue de les comparer. L'enjeu de notre recherche est principalement méthodologique : nous faisons l'hypothèse que certaines méthodes permettent d'identifier des catégories plus concrètes et pragmatiques alors que d'autres révèlent des représentations plus abstraites. Nos analyses statistiques ne nous permettent pas de vérifier cette hypothèse. Nous proposons quelques pistes pour la recherche pour discuter ces constats.

Mots-clés

Catégories d'élèves - questionnaire de jugements - descriptifs de types d'élèves - regroupements libres d'élèves.

¹ Recherche financée par le Fonds national suisse : décision 150316.

Abstract

The literature postulates that teachers rely on student categories to plan and implement their teaching. Different methods are used to examine these categories; this text investigates more precisely three of them: judgment questionnaires, descriptions of student types and free student groupings. We present these methods and a research, based on a sample of forty primary and secondary school teachers, that integrates them in order to assess the characteristics of student categories they lead to and to compare their outputs. The aim of our research is mainly methodological: we hypothesize that some methods allow to identify more concrete and pragmatic categories while others reveal more abstract conceptions. Our statistical analyzes do not match this hypothesis. We suggest some insights for further research and discuss our findings.

Keywords

Student categories - judgment questionnaire - student type descriptions - free student groupings.

1. Introduction

Pour planifier et enseigner, les enseignants recourent à des entités cognitives : croyances, connaissances, représentations, théories implicites... (Clark & Peterson, 1986 ; Crahay, Wanlin, Issaieva, & Laduron, 2010 ; Dessus, 2002). Ces entités se structurent en trois ensembles : celui de la psychologie et pédagogie générales, celui de la pédagogie de la matière enseignée et celui de la matière elle-même (Baumert & Kunter, 2006 ; Borko & Putnam, 1996 ; Shulman, 1986, 1987). Le premier ensemble nous intéresse tout particulièrement ici car il renferme les cognitions sur les élèves (Crahay *et al.*, 2010) organisées, selon la plupart des auteurs, en catégories d'élèves (*cf.* Wanlin & Laflotte, 2016). Les enseignants prennent notamment des décisions en fonction de leurs cognitions sur les élèves et de leurs analyses les concernant (Wanlin, 2009 ; Wanlin & Crahay, 2012). Ces analyses se basent sur une individualisation, un regroupement d'élèves, une mentalisation en « élève collectif » ou « steering group » (Berliner, 1987 ; Bromme, 2005 ; Dahllöf, 1971 ; Lundgren, 1973). Ces analyses, proactives et en action, reposent sur l'hypothèse d'une association, par analogie, de systèmes de cas ou catégories selon leur typicité² (Calderhead, 1996). D'autres travaux s'inspirent de la notion de « steering group » proposant la dénomination d'archiélève (Franck, 2017 ; Ronveaux, 2014). Nous ne les prenons pas en considération dans ce texte car leur prise de données ne permet pas d'identifier la nature et la structure interne de cette mentalisation d'élèves.

Notre objectif ici n'est pas d'étudier ce fonctionnement par analogies (*cf.* Wanlin & Laflotte, 2016), ni d'examiner la structure interne des catégories d'élèves (*cf.* Laflotte, Mossaz, Aliprandi, & Wanlin, 2017 ; Wanlin, Aliprandi, Mossaz, & Revilloud, 2016). Nous n'ambitionnons pas plus d'investiguer la formation ou l'effet des jugements des enseignants sur les élèves (*cf.* Bressoux, 2002 ; Bressoux & Pansu, 2003 ; Trouilloud & Sarrazin, 2003). Enfin, il ne s'agit pas non plus de confronter la notion de « steering group » à celle proprement didactique d'« archiélève » (Franck, 2015) ou de traiter des constats reposant sur des tests de performance (Maurice & Murillo, 2007).

Notre but est d'étudier les méthodes utilisées pour identifier les catégories mentales d'élèves, disponibles dans les cognitions des enseignants, plus spécifiquement celles permettant d'identifier la nature et la structure interne de ces catégories d'élèves, en se référant à la théorie du « case knowledge » de Calderhead (1996). Plus précisément, nous sondons les connaissances des enseignants sur leurs élèves *via* plusieurs approches utilisées dans la littérature sur les cognitions des enseignants sur les élèves (*cf.* Wanlin & Laflotte, 2016) pour comparer les catégories d'élèves qu'elles révèlent. Dans la foulée, nous évaluons ces méthodes et la qualité des catégories qu'elles permettent d'identifier. La première hypothèse est que certaines méthodes partant des élèves dont l'enseignant a la charge pourraient identifier des catégories plus concrètes et pragmatiques alors que les méthodes partant de représentations plus générales pourraient identifier des catégories mentales plus abstraites. Enfin, une seconde hypothèse est que les catégorisations diffèrent en fonction du degré dans lequel interviennent les enseignants. En effet, les différentes recherches citées, portant aussi bien sur les enseignants du primaire que sur les enseignants du secondaire, montrent, avec

² Dans ce texte, nous employons le terme typicité pour désigner le degré auquel un élément (ici, un élève) représente la catégorie à laquelle il est associé (représentativité) ; il s'agit de sa distance à l'exemplaire ou au prototype de la catégorie (proximité). Le degré de typicité exprime la mesure de la typicité. Pour une explicitation plus détaillée, consulter : Laflotte *et al.* (2017), Wanlin *et al.* (2016) ou Wanlin & Laflotte (2016). D'autres références en définissent les contours en psychologies sociale, cognitive et de la personnalité (Reed, 2011 ; Sternberg, 2007).

des méthodes de recueil souvent différentes, des catégories mentales non identiques d'une population à l'autre.

Nous proposons ainsi une recherche composée de trois études auxquelles ont participé 40 enseignants intervenant aux niveaux primaire ou secondaire. Après la présentation de quelques aspects théoriques, nous relatons les procédures de recueil et de traitement des données. Nous terminons par une discussion de nos résultats afin d'en dégager quelques implications pour la formation des enseignants et les recherches futures.

2. Cadre théorique

Plusieurs méthodes de recueil de données ont été utilisées par les chercheurs qui se sont intéressés aux connaissances des enseignants sur les élèves (Wanlin & Laflotte, 2016). Tous, ou presque, proposent que les connaissances sur les élèves contiennent des catégories mentalisées d'élèves qui permettent aux enseignants de fonctionner en classe et en dehors, planification comprise (Calderhead, 1983 ; Mayer & Marland, 1997). Pour révéler ces catégories mentales, ils ont recouru soit à des questionnaires (*e.g.* Hofer, 1981b) soit à des interviews durant lesquels les enseignants sont invités à réaliser des tâches de regroupement des élèves de leur classe (*e.g.* Morine-Dersheimer, 1978, 1979) ou des descriptions de types d'élèves rencontrés çà et là durant leur carrière (*e.g.* Thelen, 1967). La suite de ce texte propose d'analyser ces méthodes et leurs résultats pour dégager un plan de recherche les combinant.

2.1. Méthode 1 : Identification de catégories via questionnaire de jugements

Lecteur des travaux de Silberman (1969, 1971), Hofer (1981b, 1981a) s'insurge contre sa méthode et ses résultats. Dans ses textes, Silberman observe les traitements différentiels que les enseignants dispensent à quatre catégories d'élèves. Jusque-là, aucun problème. La critique de Hofer se situe au niveau de la provenance de la définition de ces quatre catégories : elles ont été définies *a priori* par Silberman. Or, pour Hofer, cette définition pose problème car on n'étudie pas les catégories cognitives des enseignants mais plutôt celles des chercheurs.

Fort de ses connaissances statistiques sur l'analyse en *clusters*, Hofer propose d'analyser les catégories cognitives des enseignants en collectant, par questionnaire, leurs jugements de leurs élèves émis sur plusieurs variables et que l'on identifie si ces jugements peuvent être résumés en ensembles. Pour lui, cette méthode est idoine en ce qu'elle révèle des catégories *a posteriori* d'élèves issues des cognitions enseignantes reposant sur leurs jugements.

Hofer (1981a, 1981b) identifie cinq catégories d'élèves dans les cognitions d'un échantillon d'enseignants du secondaire. Deux catégories renvoient à des élèves forts : d'une part, des élèves obtenant de bonnes notes, qualifiés d'intelligents, d'appliqués, de disciplinés, d'actifs, de travailleurs et de contrôlables. D'autre part, des élèves disciplinés obtenant aussi de bonnes notes mais qui se distinguent par leur retenue sociale : ils sont sensibles, calmes et modestes. Deux autres catégories correspondent à des élèves plus problématiques : d'un côté, des élèves peu disciplinés, perçus comme moyennement intelligents et peu enclins à travailler, de l'autre, des élèves perçus comme étant peu intéressés, peu doués, peu persévérants dans les tâches et n'ayant aucune ambition. Enfin, une dernière catégorie correspond à des élèves obtenant des scores moyens aux variables jugées par leur enseignant.

Wanlin *et al.* (2016) utilisent la même méthode auprès d'enseignants du primaire pour identifier trois ensembles d'élèves. Le premier ensemble, les *(très) bons élèves*, correspond à une réunion des deux premiers *clusters* de Hofer. Le deuxième ensemble, les *élèves en difficulté*, correspond à ses *clusters* 3 voire 5. Enfin, les *perturbateurs moyens*, peuvent être rapprochés de son quatrième *cluster*.

Cette méthode a été critiquée notamment parce qu'en utilisant une procédure d'analyse en *clusters* on identifie automatiquement des ensembles et le regroupement peut être artificiel (Oldenbürger, 1986). Pour tenir compte de cette critique, on peut compléter la méthode de Hofer avec un regroupement libre des élèves par les enseignants (*cf* méthode 3). Cette approche aboutit à des taux de recouvrement³ plutôt moyens (Hofer & Köpke, 1987 ; Wanlin *et al.*, 2016).

2.2. Méthode 2 : Identification de catégories par description de types d'élèves

Plusieurs chercheurs ont demandé à des enseignants de décrire des profils typiques d'élèves rencontrés durant leur carrière. Ces descriptions sont rassemblées en ensembles par une analyse de contenu. Thelen (1967) identifie, de cette manière, quatre ensembles auprès d'une population d'enseignants du secondaire : les bons et mauvais élèves, les élèves face auxquels les enseignants sont indifférents et ceux qu'ils estiment perdus. Kagan et Tippins (1991) constatent que des futurs enseignants du primaire décrivent les *élèves modèles* et les *dérangés peu motivés et/ou à compétences sociales moindres* ; les futurs enseignants du secondaire caractérisent les *élèves modèles non problématiques*, les *peu motivés peu préoccupants* et les *perturbateurs non scolaires*. Hörstermann, Krolak-Schwerdt et Fischbach (2010) relèvent dix *clusters* chez les enseignants du primaire : les *modèles*, les *vivants*, les « *monsieur-je-sais-tout* », les *clowns*, les *agressifs*, les *hyperactifs*, les *rêveurs*, les *fainéants*, les « *je-m'en-foutistes* » et les *retirés socialement*.

En interrogeant également des enseignants du primaire, Laflotte *et al.* (2017) mettent en évidence six *clusters* : les perturbateurs attentifs relativement bons élèves, les perturbateurs inattentifs peu performants, les calmes attentifs et attentionnés, les calmes attentifs scolaires, les moyens inattentifs introvertis à encourager et les bons élèves scolaires et attentifs. Pour évaluer cette méthode, ces chercheurs proposent de croiser ces catégories obtenues par descriptif avec celles identifiées par regroupement libre des élèves. Le taux de recouvrement entre ces deux catégorisations est moyen.

2.3. Méthode 3 : Identification de catégories par regroupements libres des élèves

Pour cette méthode, les chercheurs demandent aux enseignants de grouper soit des fiches sur lesquelles sont inscrits les prénoms et noms des élèves soit des photographies de leurs visages. Morine-Dershimer (1978, 1979) est la première à utiliser cette méthode au degré d'enseignement primaire. Hélas, elle ne décrit pas les catégories d'élèves mais s'intéresse plutôt aux variables mentionnées par les enseignants afin d'analyser si elles changent au cours de l'année scolaire. Au secondaire, Veyrac et Blanc (2014) montrent aussi que les variables

³ Il faut comprendre, par taux de recouvrement, le calcul d'un pourcentage d'élèves classés dans des catégories conceptuellement identiques obtenues par des méthodes de classification différentes. Par exemple, deux méthodes A et B obtiennent chacune 2 *clusters* conceptuellement proches 2 à 2 : $A1 \approx B2$ et $A2 \approx B1$. Tous les élèves classés en A1 par la méthode A et en B2 par la méthode B sont accolés à des ensembles conceptuellement proches donc recouvrant. Le pourcentage d'élèves classés A1-B2 et A2-B1 indiquent le taux de recouvrement des méthodes A et B.

prises en compte évoluent selon les enseignants ou le contenu qu'ils enseignent ; plus précisément, à un instant t, les chercheuses montrent des différences entre les catégorisations des débutants et des expérimentés. Enfin, elles montrent que les critères de classification peuvent évoluer en plein processus de regroupement et concernent l'engagement et la performance de l'élève dans les tâches scolaires.

Laflotte *et al.* (2017) mais aussi Wanlin *et al.* (2016) proposent une définition pour les quatre ensembles qu'ils identifient auprès des enseignants du primaire avec cette méthode : les *bons élèves peu consciencieux*, les *agités peu performants*, les *bons élèves volontaires* et les *moyens discrets peu confiants*. En comparant les ensembles ainsi obtenus avec des ensembles d'élèves obtenus *via* l'agglomérat de jugements par analyse en *clusters* et *via* l'analyse du contenu de descriptions d'élèves-types, ils montrent que, malgré des définitions catégorielles conceptuellement proches, les élèves ne sont pas fréquemment associés aux mêmes catégories selon la méthode employée. Plus encore, ils montrent que la typicité des élèves aux catégories obtenues peut être très forte pour une méthode alors que, pour l'autre, la typicité est plutôt moyenne.

2.4. Questions de recherche

De la littérature, nous retirons des catégorisations des élèves par les enseignants différentes selon la méthode employée. Nous avons des catégorisations allant de trois à dix catégories. On observe que les méthodes par agglomérat de jugements (méthode 1) ou par regroupements libres (méthode 3) identifient moins de catégories que celle qui débute par la description d'élèves typiques (méthode 2). Les études montrent aussi que le classement des élèves par une méthode ne correspond que moyennement à leur classement à partir d'une autre méthode et ce, malgré un recouvrement conceptuel fort entre les catégories identifiées. Les recherches de Laflotte *et al.* (2017) et de Wanlin *et al.* (2016) montrent que les degrés de typicité des élèves à leur catégorie diffèrent selon la méthode employée. Ainsi, la méthode par agglomérat de jugements obtient des ensembles d'élèves avec un fort degré de typicité alors que les degrés de typicité sont larges avec la méthode par analyse du contenu de descriptions d'élèves-types. Les degrés de typicité des élèves se situent entre ces deux extrêmes pour la méthode par regroupement.

Notre analyse de ces résultats nous permet de faire l'hypothèse que les méthodes identifiées dans la littérature permettent de mettre en évidence des catégories ayant un degré d'abstraction⁴ différent. Ainsi, la méthode par analyse de descriptifs (méthode 2) identifie des catégories issues des représentations profondes des enseignants ; c'est pourquoi les élèves y seraient associés à des degrés de typicité moindres. À l'inverse, la méthode par regroupements (méthode 3) permet d'identifier des ensembles pragmatiques d'élèves, utiles pour la prise de décision des enseignants ; leurs degrés de typicité devraient être plus importants. La méthode par analyse en *clusters* (méthode 1) basée sur les jugements des enseignants permettrait d'identifier des ensembles d'élèves situés entre ces deux pôles ; leurs degrés de typicité

⁴ Par degré d'abstraction nous nous référons à Calderhead (1983) qui explique que les enseignants possèdent des connaissances sur les élèves de plusieurs niveaux avec, notamment : les connaissances générales sur les enfants d'un âge donné, les connaissances sur les classes d'élèves d'un âge donné au sein de l'établissement scolaire dans lequel ils enseignent et les connaissances sur chaque élève spécifiquement. Les premières sont plus abstraites (généralement les enfants de tel âge sont capables de...) alors que les dernières sont plus concrètes (Gustave de la classe untel a des problèmes en lecture). Le degré d'abstraction renvoie à une différence de niveau cognitif théorique entre les connaissances sur les élèves en ce qu'elles sont d'ordre général ou spécifique alors que le degré de typicité renvoie à une mesure de la représentativité d'un élève particulier par rapport à une catégorie d'élèves telle que ressentie et approximée par un enseignant.

seraient néanmoins plus proches de ceux de la méthode 3. Pour vérifier cette hypothèse, nous analysons le croisement des catégorisations identifiées par ces trois méthodes. L'analyse de quelques statistiques descriptives sur les degrés de typicité des élèves permet un positionnement quant au degré d'abstraction des catégories obtenues.

Une autre hypothèse de ce texte est que les systèmes catégoriels diffèrent en fonction du degré dans lequel interviennent les enseignants. Ainsi, Kagan et Tippins (1991) montrent que les enseignants qui se préparent pour le primaire ont des conceptions plus psychologisantes que celles de leurs collègues du secondaire. Hofer (1981a) montre que les enseignants du secondaire connaissent cinq catégories d'élèves alors que leurs collègues du primaire en connaissent quatre. La comparaison des catégories en fonction du degré d'enseignement dans lequel interviennent les enseignants permet de répondre à cette question de recherche. De la même manière nous affinons l'analyse de la première hypothèse pour vérifier si les degrés de typicité des élèves obtenus *via* plusieurs méthodes diffèrent en fonction du niveau scolaire (primaire ou secondaire) dans lequel interviennent les enseignants.

Pour examiner ces questions, nous proposons une recherche composée de trois études réalisées auprès d'un échantillon de 40 enseignants.

3. Méthode

3.1. *Public*

L'échantillon est composé de 40 enseignants du canton de Genève (Suisse) dont 23 titulaires en classes de l'enseignement élémentaire et 17 en classes du secondaire I. La recherche touche 735 élèves dont 469 dans l'élémentaire (âge théorique situé entre 4 et 12 ans) et 266 dans le secondaire (âge théorique entre 12 et 16 ans). Les 40 enseignants de notre échantillon ont participé aux trois études présentées ci-dessous.

3.2. *Etudes*

La recherche est composée de trois études ayant des parties internes identiques mais des points de départ différents, et prenant appui sur chacune des trois méthodes citées dans notre cadre théorique (voir schéma 1). L'étude 1 suit le cadre théorique présenté par la méthode 1 sur la catégorisation par questionnaire de jugements. L'étude 2, quant à elle, correspond au cadre théorique décrit au sein de la méthode 2 sur la catégorisation par descriptifs de types d'élèves. Enfin, l'étude 3 se rapporte à la méthode 3 sur la catégorisation par regroupements libres des élèves.

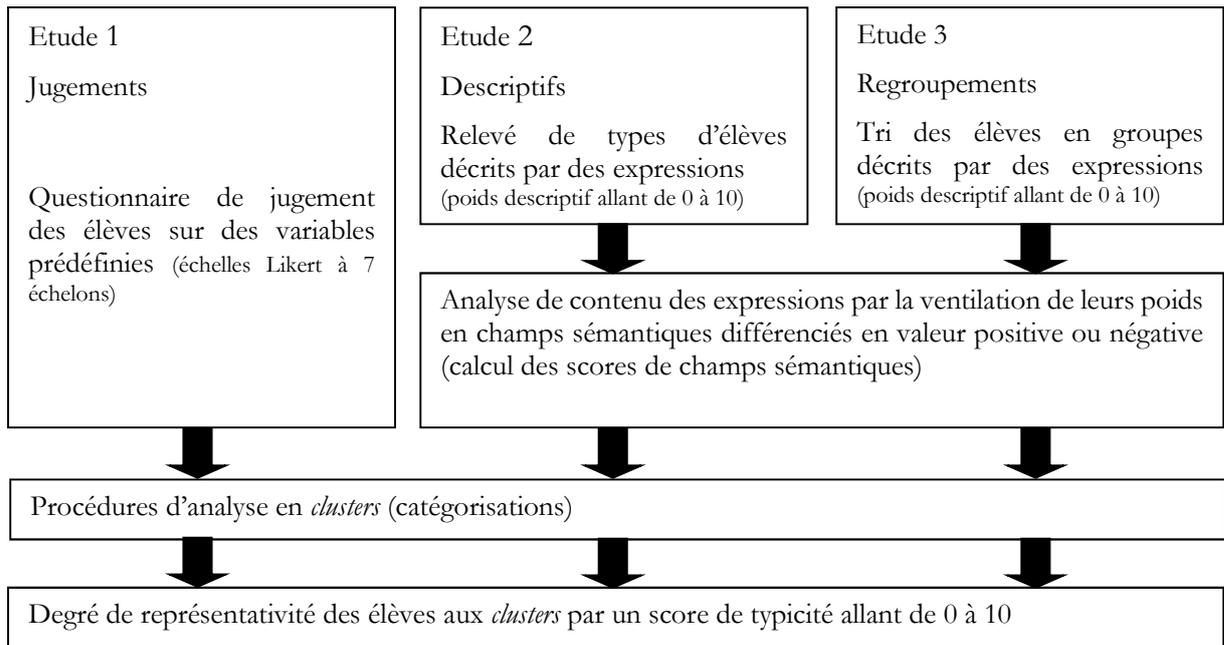


Schéma 1. Représentation du plan expérimental des trois études

3.2.1. Etude 1 : Catégories basées sur des jugements à des variables

La première catégorisation se base sur la méthode de Hofer (1981), reprise par Wanlin *et al.* (2016) : l'enseignant complète, pour chaque élève, un questionnaire grâce auquel il le juge sur une vingtaine de variables *via* un score oscillant entre 1 (pas représentatif de la variable) et 7 (fortement représentatif de la variable). Ces variables relèvent : des capacités d'apprentissage de l'élève (rétention cognitive, intelligence, ingéniosité...), de ses compétences scolaires (niveau général, niveau en mathématique et en lecture/écriture...), de ses comportements en classe (perturbations, obéissance...) et de sa personnalité (sociabilité, renfermement, timidité...)⁵.

La méthode consiste ensuite à rassembler tous les jugements émis par les enseignants sur tous leurs élèves et de les soumettre à une procédure d'analyse en *clusters* : pour Hofer (1981) une application d'au moins deux algorithmes hiérarchiques agglomératifs et, pour Wanlin *et al.* (2016), une procédure respectant les standards actuels (Clatworthy, Buick, Hankins, Weinman, & Horne, 2005 ; Everitt, Landau, Leese, & Stahl, 2011), à savoir : ajouter une vérification par un partitionnement divisif (Everitt *et al.*, 2011) et une prédiction par analyse logistique tout en appliquant une analyse de la concordance conceptuelle des catégorisations obtenues (Howell, 1998)⁶. Chaque élève est ainsi associé à un *cluster* et les comparaisons entre élèves et enseignants peuvent être effectuées.

3.2.2. Etude 2 : Descriptifs d'élèves-types associés en catégories

La deuxième catégorisation reprend la méthode initiée par Thelen (1967) et utilisée en francophonie par Laflotte, *et al.* (2017). Les enseignants décrivent des types d'élèves

⁵ La liste complète des variables mentionnées ici en quatre ensembles est disponible sur demande auprès de l'auteur de référence.

⁶ Le lecteur intéressé par les aspects statistique pourra consulter les références mentionnées (Everitt *et al.*, 2011 pour l'analyse en clusters et Howell, 1998 pour l'analyse en régression logistique ou le test d'associations) et/ou contacter le premier auteur.

différents, rencontrés durant leur carrière, et les caractérisent à l'aide d'expressions. Laflotte *et al.* (2017) demandent aux enseignants de donner, à chaque expression, un poids descriptif variant entre 1 (expression à faible valeur descriptive) et 10 (expression offrant une description parfaite du type d'élève). Cette amélioration permet, d'une part, de dépasser le comptage des expressions (Kagan & Tippins, 1991 ; Hörstermann *et al.*, 2010) et, d'autre part, d'apporter des nuances dans la procédure d'analyse en *clusters*. En effet, Hörstermann *et al.*, (2010) associent manuellement des types aux dénominations proches, comptent les expressions de ces associations puis appliquent l'analyse en *clusters*. Cette méthode a le désavantage d'aboutir à l'écartement de descriptifs aux expressions pas suffisamment proches. Laflotte *et al.* (2017) analysent le contenu des expressions classées en champs sémantiques et ventilées selon leur valeur positive ou négative ; ils retiennent le poids descriptif pour calculer un score par valeur positive et négative des champs sémantiques et leur appliquent une procédure d'analyse en *clusters* qui conserve tous les descriptifs recueillis.

Hélas, cette procédure ne permet pas d'associer les catégories de types aux élèves. C'est pourquoi, comme Laflotte *et al.* (2017), nous demandons à chaque enseignant de qualifier tous les élèves de leurs classes en donnant un score allant de 0 (pas du tout représentatif du type d'élève) à 10 (parfait représentant du type d'élève) dans chaque type d'élèves qu'ils ont décrit. Les élèves obtiennent alors l'étiquette du *cluster* de types d'élèves pour lequel ils ont le plus haut degré de typicité.

3.2.3. Etude 3 : Regroupements d'élèves associés en catégories

La troisième catégorisation reprend la méthode de Morine-Dershimer (1978, 1979). Elle consiste à demander aux enseignants de regrouper les élèves de leurs classes en ensembles. Les groupes ainsi constitués sont ensuite décrits par les enseignants et les expressions utilisées sont relevées par les chercheurs. Wanlin *et al.* (2016) et Laflotte *et al.* (2017) demandent aussi aux enseignants de donner un poids descriptif aux expressions utilisées pour procéder à une analyse du contenu en champs sémantiques selon le principe cité plus haut : constitution de scores de champs sémantiques en prenant en compte la valeur positive et négative des expressions et application d'une procédure d'analyse en *clusters* à partir de ces scores. Lors du regroupement, l'équipe genevoise demande aussi aux enseignants de donner un degré de représentativité (de 0 à 10 selon le même principe qu'expliqué dans 3.2.2.) des élèves de leur classe aux différents groupes qu'ils ont constitués pour disposer non seulement de l'appartenance des élèves aux groupes mais aussi, de leur typicité aux groupes auxquels ils sont ou non associés.

3.3. Traitement des données

3.3.1. Procédures d'analyse de contenu (études 2 et 3)

Les analyses du contenu (Bardin, 1977 ; Richards, 2005) des descriptions des types ou groupes produites par les enseignants sont menées de concert par tous les auteurs de ce texte. Concrètement, les expressions utilisées pour décrire les types ou les groupes sont classées selon leur champ sémantique. Chaque champ sémantique s'organise selon la valeur positive ou négative des expressions (Laflotte *et al.*, 2017 ; Wanlin *et al.*, 2016). Par exemple, pour le champ sémantique « comportement », nous distinguons les adjectifs « sage » (positif) et « perturbateur » (négatif) et des synonymes (calme, obéissant ou docile *vs* agitateur ou provocateur).

Chaque expression ayant obtenu un poids descriptif, nous additionnons les poids de chaque descripteur, rangés dans les pôles négatif et positif pour obtenir 10 sous-scores. Les sous-scores positifs et négatifs de chaque champ sémantique sont soustraits pour obtenir les cinq scores que nous utilisons dans les procédures d'analyse en *clusters*. La raison de la soustraction réside dans le fait que certains sous-scores obtiennent la valeur de 0 marquant l'absence d'expression descriptive; cette absence est considérée comme une donnée manquante dans le logiciel utilisé et entraîne l'ignorance du descriptif du type ou groupe d'élèves dans l'analyse. 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'entraîne pas forcément l'absence d'un sous-score négatif.

3.3.2. Procédures d'analyse en clusters (études 1, 2 et 3)

Pour les procédures d'analyse en *clusters*, nous avons respecté les standards en la matière (Clatworthy *et al.*, 2005 ; Everitt *et al.*, 2011). La procédure commence par l'utilisation des scores obtenus dans les trois études (pour l'étude 1, les scores aux variables allant de 1 à 7 et, pour les études 2 et 3, les scores de champs sémantiques). Dans SPSS 22, nous soumettons ces scores à un algorithme hiérarchique agglomératif (méthode de Ward) en utilisant le carré de la distance euclidienne comme métrique d'assemblage. Nous choisissons le nombre de *clusters* en examinant les dendrogrammes et les chaînes d'agglomération. Les *clusters* sont décrits à partir de leurs scores moyens par l'application d'une analyse de variance (Howell, 1998).

Tableau 1. Qualités statistiques des procédures d'analyse en clusters pour documenter les choix des nombres de catégories

	Etude 1	Etude 2	Etude 3
Nuées dynamiques			
- Concordance conceptuelle	OUI	OUI	OUI
- Mesure d'association	$\chi^2_{(4)} = 693,9; p < .01$	$\chi^2_{(10)} = 154,3; p < .01$	$\chi^2_{(9)} = 119,4; p < .01$
- Puissance de l'association	$V_C = 0,69; p < .01$	$V_C = 0,42; p < .01$	$V_C = 0,52; p < .01$
- Taux de recouvrement	75,8 %	59,2 %	62,8 %
Régression logistique multinomiale			
- Ajustement du modèle (Pearson)	$\chi^2_{(1404)} = 797,0; p > .05$	$\chi^2_{(82)} = 132,2; p > .05$	$\chi^2_{(288)} = 0,4; p > .05$
- Ajustement du modèle (-2LL)	$\chi^2_{(44)} = 1333,0; p < .01$	$\chi^2_{(24)} = 630,1; p < .01$	$\chi^2_{(432)} = 361,8; p > .05$
- Portion de variance expliquée (Pseudo R deux)	> 80 %	> 90 %	> 90 %
- Mesure d'association	$\chi^2_{(4)} = 1148,4; p < .01$	$\chi^2_{(10)} = 807,8; p < .01$	$\chi^2_{(9)} = 435,0; p < .01$
- Puissance de l'association	$V_C = 0,88; p < .01$	$V_C = 0,96; p < .01$	$V_C = 1,0; p < .01$
- Taux de recouvrement	92,0 %	96,8 %	100 %

Méthode split-half			
- Sous-échantillon 1			
o Mesure d'association	$\chi^2_{(4)} = 435,7 ; p < .01$	Non effectuée	Non effectuée
o Puissance de l'association	$V_C = 0,77 ; p < .01$		
o Taux de recouvrement	82,6 %		
- Sous-échantillon 2			
o Mesure d'association	$\chi^2_{(4)} = 236,8 ; p < .01$	Non effectuée	Non effectuée
o Puissance de l'association	$V_C = 0,57 ; p < .01$		
o Taux de recouvrement	56,1 %		

Le choix du nombre de *clusters* est vérifié de trois manières. D'abord, la vérification se fait par l'utilisation d'un algorithme divisif (analyse en nuées dynamiques). Ici, il n'y a pas lieu de choisir un nombre de *clusters*, le logiciel attend que son utilisateur définisse le nombre de nuées dont il suppose l'existence. Cette définition est basée sur le nombre de *clusters* identifiés lors de l'application de l'algorithme hiérarchique agglomératif. Les nuées ainsi obtenues sont aussi décrites *via* l'analyse de variance des scores. Cette description permet d'apprécier la concordance conceptuelle entre les *clusters* obtenus par les algorithmes agglomératif et divisif⁷. L'application d'une analyse de contingence (test de χ^2) et d'association (calcul du V de Cramer) permet d'apprécier la stabilité statistique des *clusters* ; autrement dit, si les deux algorithmes classent les élèves dans des catégories conceptuellement proches. Ensuite, les scores sont utilisés pour prédire, par une analyse logistique multinomiale, l'appartenance des éléments aux *clusters* auxquels ils sont associés par l'algorithme agglomératif. Enfin, seulement pour les scores de jugements⁸, l'échantillon des éléments à classer est aléatoirement découpé en deux sous-échantillons (méthode *split-half*). Les scores de chaque sous-échantillon font l'objet d'une analyse en *clusters* hiérarchique agglomérative (méthode de Ward). On vérifie alors, pour chaque sous-échantillon, si les éléments sont associés au même *cluster* que celui auquel ils ont été associés lors de l'analyse simultanée de tous les éléments de l'échantillon.

Les recouvrements entre ces trois procédures et la solution de l'algorithme hiérarchique agglomératif peuvent être représentés par un pourcentage de classification identique et des valeurs de χ^2 et de V de Cramer. Le tableau 1 montre que nos solutions agglomératives sont globalement stables d'un point de vue statistique. En effet, les valeurs des mesures d'association sont significatives et la puissance des associations est généralement forte ($> 0,50$) et les taux de recouvrement sont souvent élevés ($\geq 60\%$). Les modèles de prédiction générés par analyse logistique sont, dans l'ensemble, ajustés aux données (significativité de la statistique de Pearson $>.05$), aboutissent à des prédictions à partir des données meilleures que le hasard (significativité de la statistique $-2LL <.01$) et expliquent de larges parts de

⁷ Attention, le recouvrement conceptuel et son taux calculé, à ce stade, sont complètement internes à la méthode. Par exemple, pour la méthode 2, c'est la solution de la classification Ward qui est comparée à la classification des mêmes données par nuée dynamique. Aucune méthode n'est comparée à une autre ici.

⁸ La vérification *split-half* n'a pas été utilisée pour les scores de champs sémantiques car les éléments à classer n'étaient pas suffisamment nombreux. Ce choix repose sur les propositions d'Evrith *et al.* (2010).

variance (statistique du Pseudo R deux > 80 %)⁹. Bref, on peut accorder confiance et crédit aux catégorisations basées sur l'algorithme hiérarchique agglomératif de Ward dans nos trois études.

3.3.3. Comparaisons des clusters (études 1, 2 et 3)

Pour comparer les catégorisations de nos trois études nous analysons les données relatives aux élèves puisque nous avons, dans chaque étude, l'appartenance et le degré de typicité de chaque élève aux *clusters*. Pour l'étude 1, l'algorithme agglomératif associe chaque élève à un *cluster* et les enseignants ont apprécié la typicité de chaque élève aux *clusters* identifiés. Pour l'étude 2, les analyses portent sur le *cluster* de types pour lequel les élèves ont obtenu le plus fort degré de typicité de la part de leurs enseignants. Pour l'étude 3, les analyses retiennent les degrés de typicité des élèves au groupe dans lequel ils ont été classés.

L'appartenance des élèves aux *clusters* permet l'appréciation du recouvrement des classifications de nos trois méthodes. La typicité catégorielle des élèves permet l'évaluation des méthodes de catégorisation. En effet, une forte proximité des élèves aux *clusters* identifiés donne une indication sur la qualité du pouvoir catégoriel de la méthode. Les degrés de typicité permettent d'apprécier également si les méthodes mesurent des cognitions identiques ou si elles en dévoilent de différentes.

4. Résultats

4.1. Catégorisation selon trois méthodes de recueil de données

Cette partie du texte présente, pour nos trois études, les solutions obtenues par les procédures d'analyses en *clusters*¹⁰.

4.1.1. Etude 1 : Catégories basées sur des jugements à des variables

L'analyse en *clusters* effectuée sur les scores aux variables jugées par les enseignants identifie une solution à trois *clusters* qui peuvent être définis comme suit.

Le premier *cluster* (J1) renferme 221 élèves *négativement perçus*. Ils obtiennent des scores peu élevés aux variables relatives à la capacité de bien effectuer son métier d'élève (être attentif et concentré en classe, être appliqué lors de la réalisation des tâches ou être intéressé par la matière). Ils sont plutôt perturbateurs puisqu'ils scorent défavorablement pour les variables liées à la discipline en classe (comportement inadéquat, obéissance peu prononcée, besoin accru de recadrages disciplinaires et de valorisations). Ils ont des scores peu importants pour les variables liées à la performance scolaire (niveaux général et spécifique sous la moyenne, capacité d'apprentissage plus faible avec des besoins de guidage, de suivi et de feedbacks plus élevés). La personnalité de ces élèves est également moins bien perçue (politesse, sympathie, renfermement, équilibre, sociabilité et dignité de confiance dans la moyenne inférieure).

⁹ Le modèle prédictif de l'étude 3 sature mais nous le conservons pour trois raisons : l'analyse des dendrogrammes montre une catégorisation meilleure pour le modèle retenu, les analyses de variance avec test Scheffé distinguent clairement quatre ensembles d'élèves pour quatre scores de champs sémantiques et les modèles de régressions avec suppressions progressives des variables aboutissent à des prédictions identiques. Le lecteur intéressé peut consulter l'auteur de référence pour davantage de détails.

¹⁰ Vu la restriction en caractères, nous ne fournissons pas les dendrogrammes dans ce texte ; ils sont disponibles sur demande à l'auteur de référence.

Le troisième *cluster* (J3) renferme 225 élèves qui obtiennent des scores inverses aux élèves du *cluster* 1. Ils sont *positivement perçus* que ce soit par rapport à la perception de la part des enseignants concernant l'effectuation du métier d'élèves, la discipline ou comportement en classe, les performances et la personnalité. Les 289 élèves du deuxième *cluster* (J2) se situent entre ces deux profils. On pourrait parler d'*élèves standards*. Cette catégorisation rappelle celle de Wanlin *et al.* (2016) qui ont montré que leur système de catégories ressemble à celui de Hofer (1981a).

4.1.2. Etude 2 : Descriptifs d'élèves types associés en catégories

L'examen du dendrogramme et de la chaîne d'agglutination soutiennent le choix d'une organisation des descriptifs en cinq sous-ensembles. Dans le *cluster* T1, nous avons 32 descriptifs positivement positionnés sur l'implication (les plus positivement), le comportement et l'environnement et, dans la moyenne pour la performance, la personnalité et l'attention. Le *cluster* T3 contient 45 fiches positives en implication, performance et personnalité et, dans la moyenne en attention, comportement et environnement.

Le *cluster* T2 inclut 67 descriptifs négatifs en performance, personnalité et attention et, dans la moyenne en implication, comportement et environnement. Le *cluster* T4 englobe 25 fiches négatives en personnalité et environnement mais dans la moyenne pour les autres champs sémantiques.

Le *cluster* T5 comprend 49 descriptifs positifs en attention, dans la moyenne pour l'implication, la performance, la personnalité et l'environnement mais négatifs en comportement.

Nous identifions un système de catégories proche des précédentes études (Hörstermann *et al.*, 2010; Kagan & Tippins, 1991; Laflotte *et al.*, 2017; Thelen, 1967).

4.1.3. Etude 3 : Regroupements d'élèves associés en catégories

Les traitements statistiques retiennent quatre ensembles de groupes. Le *cluster* G1 est moins bien perçu au point de vue des performances et de la personnalité mais mieux perçu au niveau des comportements. Le *cluster* G2 est mieux perçu au niveau de l'implication et de la personnalité mais moins bien au niveau du comportement. Le *cluster* G3 obtient des scores moyens sur les variables hormis pour l'environnement où il score le plus bas. Enfin, le *cluster* G4 est positivement perçu pour les performances, l'attention et le comportement. Nous obtenons un système de catégories proche de celui des recherches précédentes (Laflotte *et al.*, 2017; Wanlin *et al.*, 2016).

4.2. Différences des catégories selon des critères spécifiques

Cette section permet d'apprécier si les catégories diffèrent selon les enseignants et le degré dans lequel ils interviennent.

Pour l'étude 1, tous les *clusters* d'élèves sont représentés chez les enseignants ($\chi^2_{(76)} = 262,74$; $p < .01$; $V_C = 0,42$; $p < .01$) mais certains identifient davantage certains types d'élèves. Les trois *clusters* apparaissent indifféremment dans les deux niveaux de la scolarité ($\chi^2_{(2)} = 6,79$; $p < .05$; $V_C = 0,10$; $p < .05$). Au primaire, 27% d'élèves catégorisés en J1, 39% en J2 et 33% en J3 ; ces pourcentages sont respectivement 35%, 39% et 26% au secondaire. La représentation des élèves au sein des *clusters* tend à être plus équilibrée au primaire qu'au secondaire où les *clusters* moins favorablement perçus sont légèrement plus représentés.

Pour l'étude 2, les enseignants ont mentionné spontanément entre 2 et 10 types d'élèves. La répartition des catégories en fonction de l'enseignant est significative ($\chi^2_{(160)} = 462,06$; $p < .001$) et d'ampleur ($V_c = 0,73$; $p < .001$). Certains *clusters* ne sont pas représentés chez certains enseignants. Ainsi, seulement 17 enseignants ont un de leurs types associé à T1 ou à T5. Pour T2, ce sont 27 enseignants qui ont au moins un de leurs types qui y est classé, 16 pour T3 et 7 pour T4. A noter que 10 enseignants voient leurs types classés dans un même *cluster* de types, 21 dans deux *clusters*, 6 dans trois *clusters* et 3 dans quatre *clusters* ; aucun enseignant ne voit de types classés dans les cinq *clusters* que nous avons identifiés. La répartition par niveau est certes significative ($\chi^2_{(160)} = 14,88$; $p < .01$) mais faible ($V_c = 0,26$; $p < .01$) ; on considèrera donc cette différence comme une tendance.

Pour l'étude 3, 19 enseignants voient leurs regroupements classés dans une même catégorie : 3 enseignants n'avaient identifié que deux groupes, 7 autres identifient trois groupes et encore 7 autres en identifient entre quatre et six. On note aussi que 22 enseignants voient leurs groupes rejoindre deux catégories alors qu'ils proposaient initialement entre deux et six groupes. Aucun enseignant n'a ses groupes rangés dans chacune des quatre catégories. A noter que la répartition des catégories en fonction du degré dans lequel enseigne l'enseignant est non significative ($\chi^2_{(3)} = 1,41$; $p = .70$; $V_c = 0,10$; $p = .70$).

Dans l'ensemble, nous avons plutôt une absence de lien entre le degré scolaire des enseignants et leurs systèmes de catégories d'élèves.

4.3. Evaluation des méthodes de catégorisation et de leur pouvoir catégoriel

L'évaluation de la capacité d'identification de systèmes de catégories à degrés d'abstraction différents est permise grâce au tableau 2. Ce tableau indique les degrés de typicité moyens des élèves aux catégories selon les méthodes utilisées et selon le niveau scolaire considéré. Il différencie les degrés de typicité moyens en fonction de la catégorie dans laquelle l'élève est rangé mais aussi ses degrés de typicité moyens aux autres catégories pour lesquelles nous ne conservons que les degrés de typicité maximal et minimal. Il propose aussi, pour chaque moyenne, l'écart-type associé ; statistique qui permet d'apprécier les dispersions des élèves autour de ces moyennes.

L'analyse des résultats ne permet pas de conclure qu'une méthode identifie des entités cognitives aux degrés d'abstraction différents. En effet, les distances moyennes des élèves à la catégorie à laquelle ils appartiennent oscillent entre 4 et 6 et les écarts-types sont relativement élevés quelle que soit la méthode employée. Leurs distances à d'autres catégories sont inférieures à 1 pour les méthodes par descriptif et regroupements et au maximum de 4 pour la méthode par jugements. Ceci ne correspond pas à notre hypothèse de degrés de typicité forts pour la méthode par regroupements pour la catégorie assignée (faible pour les autres catégories), moyens à forts pour la méthode par jugements (faibles pour les autres catégories) et faibles pour la méthode par descriptifs pour toutes les catégories.

Cette observation générale est valable quel que soit le niveau scolaire concerné. Bref, les degrés de typicité des élèves aux catégories dans lesquelles ils sont assignés sont plutôt moyens quelle que soit la méthode ; de plus, chaque méthode montre des degrés de typicité moindres pour les catégories dans lesquelles les élèves ne sont pas assignés.

Tableau 2. Degrés de typicité moyens des élèves à leurs catégories (distances) en fonction de la méthode employée : moyennes et écarts-types

Catégories	Total						Primaire						Secondaire					
	DIC		DPPAC		DPGAC		DIC		DPPAC		DPGAC		DIC		DPPAC		DPGAC	
	Moy.	E.-t.	Moy.	E.-t.	Moy.	E.-t.	Moy.	E.-t.	Moy.	E.-t.	Moy.	E.-t.	Moy.	E.-t.	Moy.	E.-t.	Moy.	E.-t.
Méthode 1 : Jugements																		
J1	5,4	3,7	2,3	3,1	2,6	3,0	2,8	3,1	3,6	3,6	4,2	3,2	8,1	1,9	1,0	1,7	1,0	1,6
J2	5,5	3,4	2,7	3,0	2,7	3,2	4,9	3,5	2,9	3,5	3,1	3,3	7,5	1,8	1,7	1,9	2,4	2,3
J3	3,9	4,0	2,8	3,5	4,0	4,0	3,0	3,9	3,0	3,8	4,6	4,3	7,4	2,1	2,4	2,4	2,4	2,2
Méthode 2 : Descriptifs																		
T1	5,8	2,0	0,2	1,2	0,6	1,4	6,5	2,0	0,0	0,0	0,8	1,4	5,3	1,9	0,2	0,7	0,8	1,7
T2	5,6	2,4	0,2	0,8	0,8	1,6	5,5	2,8	0,2	1,0	1,3	1,9	5,9	1,9	0,1	0,4	1,1	1,7
T3	5,2	2,3	0,0	0,1	0,5	1,2	5,6	2,1	0,0	0,0	0,7	1,2	4,5	2,9	0,0	0,3	1,6	1,9
T4	4,4	2,0	0,0	0,2	0,7	1,5	3,6	1,7	0,0	0,3	0,2	0,8	6,2	1,8	0,0	0,0	1,8	2,1
T5	5,8	2,0	0,0	0,0	0,7	1,6	5,8	1,9	0,0	0,0	0,6	1,5	6,1	2,1	0,0	0,0	1,5	2,2
Méthode 3 : Regroupements																		
G1	5,1	2,0	0,0	0,3	0,3	1,0	5,1	2,1	0,0	0,0	0,3	0,8	5,4	1,9	0,1	0,6	0,5	1,4
G2	5,0	2,2	0,0	0,4	0,4	1,2	6,1	2,1	0,0	0,0	0,6	1,3	4,0	1,8	0,1	0,6	0,3	1,1
G3	4,4	0,9	0,0	0,4	0,2	0,7	4,7	0,8	0,0	0,0	0,2	0,8	4,2	1,1	0,1	0,6	0,3	1,0
G4	6,3	2,4	0,1	0,7	0,6	1,8	6,3	2,3	0,0	0,0	0,2	0,7	6,5	3,0	0,7	1,3	2,2	2,6

Légende : DIC = Distance intra-catégorie ; DPPAC = Distance la plus petite à une autre catégorie ; DPGAC = Distance la plus grande à une autre catégorie ; moy. = moyenne ; E.-t. = écart-type

5. Conclusion

Notre texte poursuivait deux objectifs : (1) évaluer la capacité des méthodes à identifier des catégories d'élèves aux degrés d'abstraction différents et (2) examiner si les méthodes permettent de relever des systèmes de catégorisation différents en fonction du degré scolaire où interviennent les enseignants.

Concernant le premier, nous faisons l'hypothèse que la méthode par descriptifs (méthode 2) révèle des représentations ancrées dans les cognitions des enseignants plus éloignées des élèves composant les classes alors que la méthode basée sur un tri des élèves en groupes (méthode 3) colle plus étroitement à leurs caractéristiques. À l'entre-deux, puisque basée sur une condensation statistique de jugements, la méthode 1 identifie des catégories certes proximales à la réalité de la classe mais moins pointues que celles de la méthode 3. Nos analyses ne nous permettent pas de vérifier cette hypothèse. Pour le second objectif, nous pensions découvrir des différences : plus de catégories pour les enseignants du secondaire et une vision plus psychologisante pour les enseignants du primaire. Nos analyses statistiques ne les identifient pas.

Nos résultats défient toute logique : aucune des méthodes n'identifie au mieux les élèves dont les enseignants ont la charge vu les degrés de typicité moyens, tout comme aucune ne permet de dégager des représentations plus larges sur les apprenants. Alors que nos méthodes correspondent à celles employées en psychologie sociale ou éducationnelle, cette observation nous pousse à nous interroger sur la manière de mettre en évidence les cognitions des enseignants sur les élèves. Force est toutefois de constater que la méthode par jugements (méthode 1) permet d'obtenir des catégorisations plus stables, plus équilibrées et plus fédératrices pour les comparaisons par enseignant.

Nous conjecturons que certaines méthodes identifient des catégories pragmatiques d'élèves qui y seraient accolés avec une typicité élevée. Une des limites de notre étude est qu'elle ne permet de répondre que partiellement à cette question. En effet, l'étude des degrés de typicité est certes intéressante mais elle devrait être complétée par une analyse de la manière dont les enseignants utilisent leurs connaissances sur les élèves pour planifier et enseigner. Hofer (1981b) montre que les enseignants proposent des interactions de qualité et de quantité différentes en fonction des catégories d'élèves avec un traitement moins favorable pour les profils moins bien perçus. Par contre, Wanlin et Crahay (2011) n'observent pas ce lien. Il y a néanmoins une différence entre ces deux recherches qui utilisent les mêmes instruments : Hofer (1981b) observe un élève désigné représentatif par les enseignants par catégories alors que Wanlin et Crahay (2011) observent tous les élèves des classes. Des recherches futures devraient observer si les liens entre catégories d'appartenance des élèves et la distribution des interactions sont influencés par le degré de typicité des élèves aux catégories. L'une des contributions de notre texte est de proposer une méthode pour investiguer cette piste de recherche.

Nos résultats questionnent aussi la structure des connaissances des enseignants sur les élèves. Peut-on estimer qu'elles sont composées de catégories utiles à l'enseignement et à sa planification ? Selon Wanlin et Crahay (2012), les enseignants utilisent un ou plusieurs *steering groups* pour enseigner et Wanlin (2011) ajoute que ce sont des catégories d'élèves. Si la recherche identifie, comme nous, des catégories d'élèves, les résultats de Wanlin et Crahay (2011) questionnent en ce qu'ils montrent que les enseignants recourent plutôt à leurs sensations en situations pour cadencer les leçons. Leurs identifications des difficultés lors de

L'enseignement médiatisent les liens entre catégories d'élèves et la dispersion des interactions. Cette voie de recherche implique l'examen des variables qui interfèrent les liens entre cognitions et pratique.

Evidemment, comme toute recherche, la nôtre n'échappe pas à la critique. Certains lui reprocheront la taille de l'échantillon. Or, pour une recherche usant de méthodes qualitatives pour analyser les cognitions enseignantes, 40 enseignants est un nombre tout à fait raisonnable. Une autre critique porte sur l'étude du pragmatisme des catégorisations par des méthodes (celles de la « coutume » des recherches précédentes) postulant un accès conscient à des processus de catégorisation. Ainsi, la littérature montre que ces processus sont automatisés voire inconscients et l'emploi de ces méthodes pourrait être questionné. Peut-être que d'autres méthodes éclaireraient l'impact de l'explicitable ou du conscientisable et de ce qui ne l'est pas dans les catégories. Celles-ci pourraient, par exemple, s'appuyer sur des éléments observables et probablement non conscients : le type d'interactions reçues en classe par l'élève, la durée et la qualité des échanges, le nombre de regards obtenus... Enfin, et il s'agit là d'une remarque plus théorique, nous ne convoquons pas la théorie des stéréotypes largement étudiée en psychologie sociale (e.g. Leyens, Yzerbyt & Schadronek, 1996). La raison de ce choix réside dans l'inscription dans la théorie du « *case knowledge* » de Calderhead qui renvoie clairement à l'association de catégories sans mentionner de fonctionnement stéréotypé. Néanmoins, le lien entre catégories et stéréotypes est plausible puisque ces théories réfèrent à des notions semblables : catégories, représentations, représentativité, typicité, etc. De même, le rapprochement avec les théories des heuristiques (Shah & Oppenheimer, 2008) et de l'analogie (Sander, 2000) pourrait être intéressant. Les recherches futures pourraient recourir à ces théories ainsi qu'à leurs méthodes pour approfondir l'étude des catégories et de leurs effets sur les interactions en classe.

Mais au-delà de ces considérations, c'est pour la formation des enseignants que notre recherche peut être utile. Les enseignants ont des connaissances sur les élèves, elles contiennent des catégories d'élèves qui permettent, selon les enseignants que nous avons interviewés, de différencier leurs interventions pédagogiques. Or, le lien entre les croyances et connaissances sur les élèves et les interactions en classe n'est pas évident (Laflotte & Wanlin, 2018). Bref, si les enseignants possèdent des catégories cognitives d'élèves, les posséder n'aurait qu'un impact limité sur les interactions mais pourraient s'avérer utile pour un fonctionnement efficace en classe (planification comprise). Néanmoins, avec des degrés de typicité aussi peu élevés, c'est probablement sur la qualité des diagnostics des enseignants que la formation devrait mettre l'accent.

Il s'agit maintenant d'inviter d'autres chercheurs et formateurs à creuser ces ouvertures de recherche et de formation.

6. Références

- Bardin, L. (1977). *L'analyse de contenu*. Paris: PUF.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 4, 469-520.
- Berliner, D. (1987). Ways of thinking about students and classrooms by more and less experienced teachers. In J. Calderhead (Éd.), *Exploring teachers' thinking* (pp. 60-83). London: Cassell.
- Borko, H., & Putnam, R. (1996). Learning to teach. In D. Berliner & R. Calfee (Éds.), *Handbook of educational psychology* (pp. 673-708). New York: MacMillan.
- Bressoux, P. (2002). *Les jugements des enseignants sur la valeur scolaire des élèves* (Les stratégies de l'enseignant en situation d'interaction) (pp. 47-74). Consulté à l'adresse <http://archive-edutice.ccsd.cnrs.fr/docs/00/00/17/90/PDF/Bressoux.pdf>
- Bressoux, P., & Pansu P. (2003). *Quand les enseignants jugent leurs élèves*. Paris: PUF.
- Bromme, R. (2005). The "collective student" as the cognitive reference point of teachers' thinking about their students in the classroom. In P. Denicolo & M. Kompf (Éd.), *Teacher thinking and professional action* (pp. 31-40). London: Routledge.
- Calderhead, J. (1983). Research into teachers' and student teachers' cognitions: Exploring the nature of classroom practice. In *Annual Meeting of the American Educational Research Association*. Montreal, Canada.
- Calderhead, J. (1996). Teachers: Beliefs and knowledge. In D. Berliner & R. Calfee (Éd.), *Handbook of educational psychology* (pp. 673-708). New York: MacMillan.
- Clark, C., & Peterson P. (1986). Teachers' thought processes. In M. Wittrock (Éd.), *Handbook of research on teaching* (3d ed., pp. 255-296). New York: Mac Millan.
- Clatworthy, J., Buick, D., Hankins, M., Weinman, J., & Horne, R. (2005). The use and reporting of cluster analysis in health psychology: a review. *British Journal of Health Psychology*, 10, 329-358.
- Crahay, M., Wanlin, P., Issaieva, E., & Laduron, I. (2010). Fonctions, structuration et évolution des croyances (et connaissances) des enseignants. *Revue française de pédagogie*, 172(3), 85-129. Consulté à l'adresse <https://www.cairn.info/revue-francaise-de-pedagogie-2010-3-page-85.htm>.
- Dahllöf, U. (1971). *Ability grouping, content validity, and curriculum process analysis*. New York: Teachers College Press.
- Dessus, P. (2002). *Les effets de la planification sur l'activité de l'enseignant en classe* (Les stratégies de l'enseignant en situation d'interaction) (p. 17-28). Consulté à l'adresse <http://archive-edutice.ccsd.cnrs.fr/docs/00/00/17/90/PDF/Bressoux.pdf>
- Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis* (5th éd.). London: Wiley.
- Franck, O. (2017). *À la recherche de l'archiélève lecteur à travers l'analyse du geste de planification: rôle des élèves dans les modifications de séquences d'enseignement*. Thèse de doctorat en sciences de l'éducation, Université de Genève, Suisse.
- Franck, O. (2015). L'archiélève lecteur dans le discours de planification d'enseignants du primaire, du secondaire I et du secondaire II: effet de deux textes littéraires contrastés. *Le point sur l'enseignement et l'apprentissage de la lecture du primaire au postsecondaire*. Présenté à 14^e Rencontres internationales du réseau de Recherche Education et Formation, Montréal.
- Hofer, M. (1981a). Die Schülerspezifität in Einstellungen und Verhaltensweisen des Lehrers. In M. Haidl (Éd.), *Lehrerpersönlichkeit und Lehrerrolle im sozial-integrativen Unterricht* (pp. 58-82). München: Lurz.

- Hofer, M. (1981b). Schülergruppierungen in Urteil und Verhalten des Lehrers. In M. Hofer (Éd.), *Informationsverarbeitung und Entscheidungsverhalten von Lehrern* (pp. 192-221). München: Urban & Schwarzenberg.
- 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.
- Howell, D. (1998). *Méthodes statistiques en sciences humaines*. Bruxelles: De Boeck.
- Kagan, D., & Tippins, D. (1991). How student teachers describe their pupils. *Teaching and Teacher Education*, 7(5/6), 455-466.
- Laflotte, L., Mossaz, A., Aliprandi, M.-L., & Wanlin, P. (2017). Catégorisation des apprenants: une étude de la structure interne des cognitions des enseignants sur les élèves. *Revue suisse des sciences de l'éducation*, 39(1), 147-169.
- Laflotte, L., & Wanlin, P. (2018). Approche typologique des connaissances d'enseignants en formation pour le secondaire sur les élèves et leurs caractéristiques motivationnelles. *Formation et pratiques d'enseignement en questions*, 23, 195-212. Consulté à l'adresse http://www.revuedeshep.ch/site-fpeq-n/Site_FPEQ/23_files/23-13-Laflotte.pdf
- Leyens, J.-P., Yzerbyt, V., & Schadron, G. (1996). *Stéréotypes et cognition sociale. Etude des processus mentaux en psychologie sociale*. Liège: Mardaga.
- Lundgren, U. (1973). Pedagogical frames and the teaching process. A report from an empirical curriculum project. In *Annual Meeting of the American Educational Research Association*. New Orleans, Louisiana.
- Mayer, D., & Marland, P. (1997). Teachers' knowledge of students: a significant domain of practical knowledge? *Asia-Pacific Journal of Teacher Education*, 25(1), 17-34.
- Maurice, J. J., & Murillo, A. (2007). *La distance à la performance attendue, un indicateur prévisionnel de l'efficacité des enseignants - en lecture, au cours préparatoire*. Présenté à Les effets des pratiques enseignantes sur les apprentissages des élèves, Besançon.
- Morine-Dersheimer, G. (1978). How teachers "see" their pupils. *Educational Research Quarterly*, 3(4), 43-52.
- Morine-Dersheimer, G. (1979). *Teachers' conceptions of pupils - An outgrowth of instructional context: The south bay study, part III* (Research Series No. 59) (p. 67). Washington, DC: Michigan State University, Institute for Research on Teaching.
- Oldenbürger, H. (1986). Does a tendency to group pupils on attributes exist in teachers' cognitions/judgements? In M. Ben Peretz, R. Bromme & R. Halkes (Éd.), *Advances of research on teacher thinking* (p. 186-200). Lisse: Swets & Zeitlinger B.V.
- Reed, S. (2011). *Cognition : Théories et applications*. Bruxelles: De Boeck.
- Richards, L. (2005). *Handling qualitative data: a practical guide*. London: Sage.
- Ronveaux, C. (2014). L'archiélève lecteur entre tâche, activité et performance de lecture. In J.-L. Dufays & B. Daunay (Éds.), *Didactique du français langue première : quelle place pour le point de vue des élèves ?* (pp. 119-138). Bruxelles: De Boeck.
- Sander, E. (2000). *L'analogie, du naïf au créatif : Analogie et catégorisation*. Paris: L'Harmattan.
- Shah, A., & Oppenheimer, D. (2008). Heuristics made easy: an effort-reduction framework. *Psychological Bulletin*, 134(2), 207-222.

- Shulman, L. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14.
- Shulman, L. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educational Review*, 57(1), 1-22.
- Silberman, M. (1969). Behavioral expression of teacher's attitudes toward elementary school students. *Journal of Educational Psychology*, 60, 402-407.
- Silberman, M. (1971). Teacher's attitudes and actions towards students. In M. Silberman (Éd.), *The experience of schooling*. New York: Holt.
- Sternberg, R. (2007). *Manuel de psychologie cognitive: du laboratoire à la vie quotidienne*. Bruxelles: De Boeck.
- Thelen, H. (1967). *Classroom grouping for teachability*. New York: John Wiley & sons, inc.
- Trouilloud, D., & Sarrazin, P. (2003). Les connaissances actuelles sur l'effet Pygmalion : processus, poids et modulateurs. *Revue Française de Pédagogie*, 145, 89-119.
- Veyrac, H., & Blanc, J. (2014). *Étude de la catégorisation des élèves par les professeurs*, Rapport final Convention d'études et de recherche Direction de l'évaluation, de la prospective et de la performance du Ministère de l'éducation nationale (France), Commissariat général à l'égalité des territoires, Le Défenseur des Droits, Appel d'offre « égalité des chances à l'école », Université de Toulouse.
- Wanlin, P. (2009). La pensée des enseignants lors de la planification de leur enseignement. *Revue Française de Pédagogie*, 166, 89-128.
- Wanlin, P. (2011). *Elèves forts ou faibles : qui donne le tempo ? Une analyse de la place des élèves dans les processus de pensée des enseignants*. Thèse de doctorat, Université de Liège, Belgique.
- 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.
- 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.
- Wanlin, P., & Crahay, M. (2012). La pensée des enseignants pendant l'interaction en classe. *Éducation et Didactique*, 6(1), 9-46.
- Wanlin, P., & Laflotte, L. (2016). *Connaissances des enseignants sur le(ur)s élèves et leur utilisation pour donner cours* (FNS 150316). Genève : Université de Genève, archive ouverte. Consulté à l'adresse <https://archive-ouverte.unige.ch/unige:94562/>