

---

# Élaboration et mise en œuvre d'un dispositif *ad hoc* d'évaluation de politique publique : le cas concret de l'évaluation de la politique de dédoublement des classes de CP et de CE1 en zones d'éducation prioritaire (2017-2021)

Ronan VOURC'H (\*), Axelle CHARPENTIER (\*), Fabrice MURAT (\*), Thierry ROCHER (\*)

(\*) Direction de l'évaluation, de la prospective et de la performance, MENJS  
ronan.vourch@education.gouv.fr

**Mots-clés.** : Évaluation multidimensionnelle, collecte, impact, politique publique, éducation.

**Domaines.** Collecte de données d'enquêtes. Évaluation des politiques publiques

---

## Résumé

Cet article porte sur l'évaluation d'impact de la mesure de réduction de la taille des classes de CP et de CE1 en REP+. Cette évaluation a été mise en œuvre par la Direction de l'évaluation, de la prospective et de la performance (DEPP) du Ministère de l'Éducation nationale, de la Jeunesse et des Sports, à partir de la rentrée 2017. L'étude de la DEPP propose plusieurs contributions à la littérature scientifique sur la réduction de la taille de classe : d'une part, elle offre une mesure des effets sur la performance scolaire dans le contexte d'une politique publique implémentée à très grande échelle ; d'autre part, l'évaluation explore les processus à l'œuvre à partir de données d'enquête et d'observation portant à la fois sur les pratiques d'enseignement et sur le comportement des élèves en classe.

Le protocole d'observation et d'évaluation statistique a été conçu, en amont de la politique évaluée, par la DEPP avec l'appui de chercheurs spécialistes de l'évaluation des politiques publiques et des sciences de l'éducation. Les collectes de données conduites auprès des élèves, des directeurs d'école et des enseignants à partir de la rentrée scolaire 2017 ont permis de mesurer l'impact de la réduction de la taille de classe à court et moyen terme sur la progression des élèves en REP+ (scolarisés en CP en 2017-2018, puis en CE1 en 2018-2019), de décrire les modalités de déploiement de la mesure, de recueillir des informations sur les pratiques d'enseignement en classe de CP puis de CE1, et d'observer *in situ* l'engagement des élèves, les interactions professeur-élèves ainsi que les pratiques pédagogiques en classe de CP.

# Abstract

This article focuses on the multidimensional approach used by the Directorate for Evaluation, Forecasting and Performance monitoring (DEPP) in collaboration with researchers to evaluate the impact of a class size reduction policy in first and second grades in high priority educational areas (REP+), implemented from the start of the 2017 school year. It offers several contributions to the scientific literature on class size reduction : on the one hand, it provides a measure of the effects on school performance in the context of a public policy implemented on a very large scale ; on the other hand, the study explores the processes at work on the basis of survey and observation data on both teaching practices and the behaviour of pupils in class.

## Introduction

Pour combattre la difficulté scolaire dès les premières années des apprentissages fondamentaux et soutenir les élèves les plus fragiles, une mesure de réduction importante de la taille des classes de CP en éducation prioritaire a été mise en place à partir de la rentrée 2017.<sup>1</sup>

Son déploiement a été étalé sur trois années scolaires successives :

- À la rentrée 2017 : réduction de la taille des classes de CP en REP+ (Réseau d'éducation prioritaire renforcé), visant un effectif de 12 élèves, soit presque un dédoublement par rapport aux effectifs de REP+ des années antérieures et à ceux hors éducation prioritaire.
- À la rentrée 2018 : extension de la mesure aux élèves de CE1 en REP+ (donc généralement ceux ayant déjà bénéficié de la mesure en CP) et aux élèves de CP en REP (Réseau d'éducation prioritaire).
- À la rentrée 2019 : extension de la mesure aux élèves de CE1 en REP.

L'objectif du présent article est de donner à voir l'élaboration et la mise en oeuvre du dispositif *ad hoc* d'évaluation de la DEPP (Direction de l'évaluation, de la prospective et de la performance). La Section 1 présente une revue de la littérature scientifique relative à l'évaluation des politiques de réduction de taille de classe ayant nourri la conception du protocole d'évaluation multidimensionnelle dont les finalités et la méthodologie d'échantillonnage sont décrites à grands traits dans la Section 2. La DEPP a mobilisé tant des données administratives que des données d'enquête *ad hoc* pour rendre compte finement des modalités de mise en oeuvre de la réforme (Section 3). La mesure de la progression des résultats des élèves bénéficiaires est décrite et illustrée dans la Section 4. Enfin, les Sections 5 et 6 présentent quant à elles les protocoles d'enquête et d'observation de classe mis en place afin d'apprécier l'évolution des conditions d'enseignement et d'apprentissage. Tout au long de l'article, certains résultats d'évaluation sont fournis à titre d'illustration ; le lecteur intéressé par l'ensemble des résultats de cette évaluation de politique publique pourra se référer à l'étude complète de la DEPP (Andreu *et al.*, 2021<sup>2</sup>).

---

1. Rappelons que depuis la rentrée 2015 (faisant suite aux ZEP, RAR, ECLAIR), l'éducation prioritaire est actuellement structurée en deux dispositifs : REP+ (Réseau d'éducation prioritaire renforcé) et REP (Réseau d'éducation prioritaire renforcé). Les collèges accueillant des élèves socialement défavorisés (enfants d'ouvriers, boursiers) ont intégré ces dispositifs, les plus défavorisés se trouvant en REP+. Les écoles alimentant ces collèges ont été aussi intégrées pour constituer un réseau cohérent d'action pédagogique. Cela signifie que si les collèges en REP+ se distinguent très nettement des collèges en REP et encore plus hors éducation prioritaire, il y a moins de différences lorsqu'on compare les écoles, le recouvrement entre REP+ et REP ayant permis de construire un groupe témoin d'écoles REP comparables aux écoles REP+. À la rentrée 2020, 2 458 écoles étaient classées en REP+, scolarisant un peu plus de 279 000 élèves en élémentaire (soit environ 8 % de la population). La part d'élèves en REP est de 12 %.

2. Les auteurs du présent article remercient très chaleureusement, pour leurs conseils précieux tout au long de la mise en oeuvre de l'évaluation, les membres de son comité scientifique : Pascal Bressoux (Université Grenoble-Alpes), Marc Gurgand et Julien Grenet (École d'économie de Paris). Ils associent

# 1 Impact de la réduction de la taille des classes : que dit la littérature scientifique ?

La mise en place de la réduction de la taille des classes de CP en éducation prioritaire s’est appuyée sur des travaux de recherche, synthétisés dans cette partie, qui mettent en évidence des effets importants de cette mesure sur les résultats des élèves, avec un effet plus marqué pour les élèves défavorisés et maximal dans le premier degré.

## 1.1 Des effets généralement positifs de la réduction de la taille des classes

Plusieurs revues de littérature récentes sur la taille de classe (par exemple, Bouguen, Grenet & Gurgand, 2017 ; Monso, 2014) s’appuyant sur des résultats de recherche parmi les plus robustes méthodologiquement montrent qu’au primaire, la réduction du nombre d’élèves dans les classes a un effet bénéfique sur les performances des élèves. L’un des travaux phares de la littérature mobilise les données de l’expérimentation à grande échelle STAR (Student-Teacher Achievement Ratio) menée à partir de 1985 dans des écoles primaires de l’État du Tennessee. Dans ce cadre, des élèves et enseignants ont été aléatoirement assignés à des tailles de classe à effectif réduit (13 à 17 élèves par enseignant) ou à effectif normal (22 à 25 élèves par enseignant). Ce protocole d’évaluation dite contrôlée (ou aléatoire) permet d’attribuer avec une plus grande confiance les résultats observés sur les apprentissages à la réduction de la taille de classe, car le tirage au sort garantit que les résultats ne soient pas tirés par des différences initiales entre les élèves ou des différences entre les enseignants. Le projet STAR a ciblé les élèves dès leur entrée en grande section de maternelle et s’est poursuivie jusqu’en troisième année de scolarité en élémentaire (l’équivalent du CE2 en France). Le protocole d’évaluation a suivi près de 12 000 élèves pendant quatre ans. Pendant toute la durée de l’expérimentation (entre 1985 et 1989), les élèves devaient rester dans le même type de classe auquel le tirage au sort les avait alloués initialement. Les évaluations du projet STAR (Krueger, 1999 ; Krueger & Whitmore, 2001 ; Chetty et al. 2011) ont montré que les classes à effectif réduit ont un impact positif non seulement sur les apprentissages et compétences des élèves mesurés pendant la durée de l’expérimentation, mais également à plus long terme sur des indicateurs mesurés plusieurs années après la fin de l’expérimentation (par exemple, l’obtention du diplôme d’études secondaires, l’accès à l’université et l’obtention d’un diplôme d’études supérieures, l’accès à la propriété, le comportement d’épargne ou le fait de se marier).

Dans un autre contexte, en Suède, Fredriksson et al. (2013) identifient également des effets à long terme de la réduction de la taille des classes à l’école primaire. Pour isoler l’impact de la politique évaluée, ils exploitent la variation exogène de la taille des classes créée par une règle de taille maximale des classes. Leurs résultats montrent que des classes à effectif réduit au cours des trois dernières années de l’école primaire (de 10 à 13 ans), pour des cohortes d’élèves nés entre 1967 et 1982, sont bénéfiques non seulement pour les résultats à des tests de compétences verbales et de logique à l’âge de 13 ans, mais aussi pour les résultats à des mesures des compétences socio-émotionnelles (estime de soi et persévérance notamment) au même âge, pour les résultats aux tests cognitifs ultérieurs (aux âges de 16 et 18 ans), ainsi que pour le fait d’avoir terminé ses études et pour le niveau des salaires observés entre 27 à 42 ans. La méthodologie d’évaluation reposant sur la comparaison des résultats des élèves des classes dont les effectifs sont juste en dessous et juste au-dessus des seuils d’ouverture de classes, utilisée par Fredriksson et al. (2013),

---

par ailleurs à cette contribution leurs collègues des équipes de la DEPP ainsi que les chercheurs du LaRAC impliqués dans la mise en œuvre et les analyses des collectes de données décrites dans cette contribution, qui s’appuie largement sur les éléments présentés par Andreu *et al.* (2021).

est également celle mise en œuvre par Angrist & Lavy (1999) sur des données israéliennes. Leur étude montre un effet positif de la réduction de la taille de classe sur les compétences en lecture en CM2 (mais pas en CM1, niveau qui a également été testé).

En France, en dépit de quelques résultats non significatifs (Schmitt-Roland & Thuarel-Richard, 1997; Verdon & Thuarel-Richard, 1997), les résultats obtenus confirment largement la tendance internationale, qu'il s'agisse de travaux expérimentaux (Bressoux & Lima, 2011), de travaux utilisant des variables instrumentales (Piketty, 2004; Piketty & Valdenaire, 2006) ou encore de travaux corrélatifs (Bressoux, 1996; Bressoux, Kramarz & Prost, 2009; Mingat, 1991).

Pris dans leur ensemble, les travaux d'évaluation les plus robustes n'apportent pas de réponse tranchée quant aux effets attendus si la politique de réduction de taille de classe est ciblée sur les écoles concentrant la difficulté scolaire, à la fois parce que la question de l'hétérogénéité des effets n'a pas été explorée de manière systématique dans les travaux recensés et parce que lorsque cela a été fait, les résultats ne convergent pas tous (Bouguen et al., 2017). Il existe toutefois un faisceau d'évidences (par exemple, Krueger, 1999; Piketty & Valdenaire, 2006) suggérant que les effets bénéfiques pourraient être plus forts pour les élèves d'origine sociale défavorisée ou appartenant à des minorités.

## 1.2 Mais des débats concernant les politiques à mener

Pourtant, en dépit de l'orientation générale des résultats, il demeure beaucoup de controverses entre les auteurs sur le bien-fondé de politiques de réduction de la taille des classes. Tout d'abord, parce que toutes les études ne montrent pas d'effet statistiquement significatif. Par exemple, l'expérience en Californie qui a tenté de répliquer l'étude STAR faite dans le Tennessee n'a pas montré de résultats probants (Bohrnstedt, Stecher & Wiley, 2000)<sup>3</sup>. Un article récent de Angrist et al. (2019) exploitant les seuils d'ouverture de classes sur des données israéliennes ne retrouve pas l'effet identifié dans Angrist & Lavy (1999). Ensuite, parce que des institutions influentes, telle l'OCDE, soulignent que, dans leurs études internationales, certains pays très performants (d'Extrême-Orient notamment) affichent des tailles de classes élevées (bien que ces résultats ne soient que corrélationnels). Ces institutions affirment également que la qualité de l'enseignement joue bien plus que la taille de la classe (O'Malley & Schleicher, 2014, cités par Hattie, 2016). Enfin et surtout parce que certains remettent en cause non pas le potentiel effet positif d'une réduction de la taille des classes, mais son rapport coût-efficacité : réduire la taille des classes coûte cher car il faut embaucher davantage d'enseignants, or certains auteurs arguent que d'autres mesures, moins coûteuses, telles que la formation des enseignants à des pratiques efficaces, produisent des effets bien plus importants pour un coût moindre (Hattie, 2016).

D'autres auteurs répliquent toutefois qu'il n'est guère pertinent de comparer les bénéfices de la réduction de la taille des classes avec ceux fournis par les pratiques d'enseignement, car la variable « taille des classes » ne fait qu'identifier un élément du contexte d'apprentissage : elle ne décrit pas en soi les pratiques et les processus à l'œuvre en classe comme peuvent le faire des variables qui décrivent spécifiquement l'organisation, les interactions ou la coopération en classe, par exemple. La taille de classe n'est que le contexte à partir duquel un grand nombre de ces pratiques et processus peuvent être mis en œuvre. En elle-même, la variable « taille de classe » ne dit pas quelles pratiques d'enseignement sont les plus appropriées selon les variations de taille de classe (Blatchford, 2016). Par ailleurs, la taille de classe est une caractéristique facile à changer et sans équivoque, pour peu qu'on en ait la volonté politique et les moyens bien sûr. Au

---

3. Une des raisons à ces résultats décevants semble toutefois tenir dans le fait que cette vaste expérience a été menée à la hâte et qu'elle a conduit à recruter pour les classes réduites de nouveaux enseignants mal, voire non formés (Bohrnstedt, Stecher & Wiley, 2000; Jepsen et Rivkin, 2009; Hattie, 2016).

contraire, changer les pratiques enseignantes est très difficile et ce d'autant plus quand il s'agit de réaliser ce changement à grande échelle. Même si la mise en œuvre de pratiques efficaces, telles que révélées par la littérature scientifique (e.g. Hattie, 2009), pourrait en théorie produire plus de gains que la réduction de la taille des classes – ce que peu de chercheurs contestent –, ces arguments demeurent néanmoins théoriques, car les études montrent qu'il existe de nombreux obstacles à l'appropriation par les enseignants de ces pratiques prometteuses, ce surtout lorsqu'on envisage une transformation des pratiques à large échelle (Bryk, 2014).

### 1.3 La question des tailles d'effets

L'ampleur des tailles d'effet rapportées concernant les expériences de réduction substantielle de la taille des classes se situe généralement aux alentours d'une amélioration de la performance scolaire de l'ordre de 20 % à 30 % d'un écart-type (Bouguen et al., 2017). Dans une classe de 24 élèves, l'effet du dédoublement permettrait donc à l'élève médian de progresser de 2 à 3 rangs dans la classe. Bouguen et al. (2017) précisent que cette taille d'effet « est d'un ordre de grandeur comparable à la moitié de l'écart de performance moyen que l'on observe à l'entrée en CP entre les enfants de PCS favorisées (cadres, professions intellectuelles supérieures) et les enfants de PCS défavorisées (ouvriers, personnes sans activité professionnelle) ». Une autre manière de rapporter les tailles d'effet consiste à les calculer pour chaque élève en moins dans la classe : les effets estimés, pour chaque élève en moins dans la classe, sont alors compris entre 1,7 % et 2,5 % d'un écart-type.

Cette taille d'effet est considérée comme modeste par certains auteurs (e.g. Hattie, 2009) et elle est bien sûr au cœur du débat sur le rapport coût-efficacité des mesures de réduction de la taille des classes. Il n'est cependant pas inutile de placer cette taille d'effet en comparaison avec d'autres expériences pédagogiques randomisées à grande échelle. Dans un article récent, Kraft (2020) rapporte que les expériences randomisées<sup>4</sup> à grande échelle conduisent à des effets notoirement plus faibles que les autres études qui, de par leurs petites tailles d'échantillon, et de par la possibilité d'effets confondus, sont moins fiables. Kraft parvient à un effet moyen de 6 % d'un écart-type, qui se révèle donc globalement faible. Ne prenant en compte que les études randomisées, il est important de relever que Kraft qualifie donc ainsi les effets causaux des opérations. Kraft déduit de sa synthèse que, compte tenu de la distribution des effets observés, de nouveaux benchmarks doivent être établis concernant la force des effets de ce type d'expériences. Selon lui, un effet de 5 % à 20 % d'un écart-type devrait être considéré comme « moyen » (*medium*) tandis qu'un effet supérieur à 20 % d'un écart-type serait qualifié de « fort » (*large*). On voit donc qu'un effet de 20 % à 30 % d'un écart-type de la réduction de la taille des classes tel qu'indiqué supra serait loin d'être anodin puisqu'à la frontière entre un effet moyen et un effet fort<sup>5</sup>.

### 1.4 La question des processus à l'œuvre

Bien qu'on dispose de nombreux résultats qui viennent étayer cet effet bénéfique de la réduction de la taille des classes en début de scolarité, peu de résultats sont disponibles quant aux

---

4. Les expériences randomisées (appelées aussi « expériences aléatoires » ou « expériences contrôlées ») consistent à comparer deux groupes formés aléatoirement : l'un est bénéficiaire d'un dispositif dont on cherche à évaluer les effets, l'autre ne l'est pas (il sert de groupe de comparaison). Sous certaines conditions, ce type de protocole fournit un niveau de preuve supérieur à celui d'autres méthodologies quantitatives. Voir par exemple Gurgand & Valdenaire (2012) et Fougère (2012) pour une discussion des méthodes.

5. Kraft prend aussi en compte le coût et il faudrait donc estimer, pour chaque expérience le coût par élève afin d'évaluer le rapport coût-efficacité de l'opération

processus qui permettent d'obtenir ce bénéfice. C'est ce que soulignent Harfitt et Tsui (2015) lorsqu'ils font le constat que la façon dont la taille des classes influence les processus d'enseignement-apprentissage est une question encore peu explorée et largement débattue.

La plupart des études qui ont cherché à étudier l'impact de la réduction de la taille des classes sur les pratiques d'enseignement-apprentissage et le fonctionnement de la classe ont utilisé des protocoles fondés sur l'interrogation des enseignants. Il existe donc des résultats sur la façon dont les enseignants perçoivent la réduction de la taille des classes dans lesquelles ils enseignent. Comme le rappelle Harfitt (2012), les enseignants qui enseignent dans des classes à effectif réduit ont très généralement une attitude positive face à cette réduction. Par rapport à leur propre pratique dans les classes de taille « standard », ils indiquent percevoir que les comportements des élèves sont plus adaptés au contexte scolaire, que la gestion de la classe est plus simple et que l'attitude des apprenants est plus positive, qu'ils consacrent plus de temps à l'enseignement et sont plus susceptibles d'individualiser l'enseignement, ce d'autant plus qu'ils déclarent mieux connaître chaque élève. Ainsi, Galton et Pell (2009) concluent que les enseignants font état de moins de stress et d'un confort professionnel accru lorsqu'ils enseignent dans des classes à effectif réduit.

Toutefois, les observations menées en classe ne confirment pas toujours les représentations des enseignants. La plupart des études observationnelles montrent que les pratiques enseignantes changent assez peu dans les classes réduites : les enseignants ne changent pas automatiquement leurs approches pédagogiques et ils n'individualisent pas forcément plus leur enseignement (Finn & Shanahan, 2016). Pourtant, cela ne signifie pas que la réduction de la taille des classes n'ait pas d'effets : d'après Finn et Shanahan, les petites classes produisent des bénéfices « en elles-mêmes et par elles-mêmes » (2016).

Selon Finn et al. (2003) un facteur important qui différencierait les classes à effectif réduit des autres classes serait le temps d'engagement des élèves dans les tâches scolaires, qui serait accru dans les classes à effectif réduit. Ils expliquent ce fait par une « visibilité » accrue des élèves, qui limite les possibilités de « paresse sociale » (*social loafing*) et conduit donc à davantage de temps effectif de travail en classe. Par exemple, dans une étude conduite en Virginie et en Californie, rapportée par Finn et al (2003), le temps d'engagement moyen des élèves dans les tâches scolaires était de 56,6 % en contexte de grande classe, alors qu'il était de 74,6 % en contexte de petite classe. Concernant le temps où les élèves attendent en n'ayant rien à faire (down time), il était en moyenne de 13,9 % en contexte de grande classe, mais de 0 % en contexte de petite classe. Ces résultats permettent au moins en partie d'expliquer le bénéfice enregistré sur les apprentissages, les élèves des classes à effectif réduit passant plus de temps sur les tâches scolaires que les élèves des autres classes. Toutefois, si dans leur vaste étude CSPAR menée en Grande-Bretagne, Blatchford, Bassett et Brown (2005) ont trouvé eux aussi un lien entre engagement et taille de la classe en début de scolarité, ils n'ont pas retrouvé ce lien en fin de primaire.

Une autre caractéristique des classes à effectifs réduits issue d'observations in situ, est l'augmentation du nombre d'interactions individuelles enseignant-élève portant sur les contenus d'apprentissage et initiées par un élève (Blatchford et al., 2005). Des résultats similaires ont été observés en début de scolarité par Fomer-Annevelink et al. (2010) qui indiquent que les élèves bénéficient de plus d'interactions individuelles élève-enseignant dans les classes à effectif réduit (15 élèves) que dans les classes à effectif habituel (24 élèves), tant en ce qui concerne les interactions portant sur les contenus qu'en ce qui concerne celles liées à la gestion de la classe, avec des tailles d'effet respectives au CP de 1,04 et de 0,43 écart-type, ce qui peut être considéré comme très fort. De plus, dans cette étude, tous les élèves, quel que soit leur niveau d'habileté ou leur comportement, bénéficiaient de la même façon de cette augmentation des interactions

individuelles.

Bien que d'autres études observationnelles ne retrouvent pas ce résultat (par exemple, Evertson & Folger, 1989), Buckingham (2003) a estimé que, dans une séance de 40 minutes dont la moitié serait organisée comme une leçon collective et l'autre comme un temps de travail individuel pendant lequel l'enseignant interagirait individuellement avec chaque élève en répartissant son temps équitablement entre les élèves, chaque élève pourrait bénéficier de 30 secondes d'attention de la part de l'enseignant dans une classe de 40 élèves. Ce temps augmenterait mécaniquement lorsque le nombre d'élèves présents diminue pour passer à 48 secondes avec 25 élèves, à 1 minute avec 20 élèves, 1 minute 20 secondes avec 15 élèves et 2 minutes avec 10 élèves. Si on considère qu'il y a 5 séances d'apprentissage dans une journée scolaire, par rapport à une classe de 40 élèves, un élève dans une classe de 20 élèves bénéficierait de 2 minutes 30 d'attention supplémentaire de la part de l'enseignant chaque jour de classe. De même, pour un passage de 20 à 10 élèves, le gain serait de 5 minutes. Pour une année scolaire comprenant 180 jours, cela représenterait un écart de 7 h 30 dans le cas d'un passage de 40 à 20 élèves, de 15 h dans le cas d'un passage de 20 à 10 élèves. Bien qu'il s'agisse là d'une estimation théorique, cela donne une illustration convaincante de ce qu'entraîne « en elle-même et par elle-même » la réduction de la taille des classes.

Au-delà de ce simple effet « mécanique » de la réduction de la taille des classes, il est toutefois probable que le bénéfice engendré puisse lui-même être amplifié par des pratiques pédagogiques adaptées. C'est notamment la thèse que défendent Blatchford et Russell (2020). Selon eux, on ne peut se contenter du fait que la réduction de la taille des classes produit en elle-même et par elle-même des effets bénéfiques. Il faut comprendre comment la taille des classes s'interconnecte avec les processus en classe, ce qu'ils ambitionnent de faire en développant une « approche sociopédagogique » (*social pedagogical approach*). Les effets de la réduction de la taille des classes pourraient être notablement augmentés si les enseignants tiraient, dans leurs pratiques d'enseignement, le meilleur parti de la réduction de la taille des classes.

Ainsi, il est évident que le nombre d'interactions individuelles enseignant-élève n'augmentera sensiblement avec la réduction de la taille des classes que dans la mesure où les pratiques pédagogiques adoptées génèrent elles-mêmes déjà un certain nombre d'interactions de ce type, y compris en contexte de classe nombreuse. L'enjeu pourrait être alors d'inciter les enseignants à mettre en place une organisation pédagogique qui favorise des interactions individuelles enseignant-élève nombreuses et de qualité.

## 2 Protocole d'évaluation

En juin 2017, la DEPP a été chargée, au travers d'évaluations des compétences des élèves et d'observations des pratiques enseignantes sur échantillons, d'évaluer la réduction de la taille des classes de CP et de CE1 en REP+, dispositif qui sera appelé dans la suite CP12, car le premier objectif affiché était d'atteindre un effectif de 12 élèves en CP en REP+. Pour ce faire, elle a constitué, dès la phase de conception du protocole, un groupe consultatif de chercheurs de différentes écoles et universités : Marc Gurgand (École d'économie de Paris), Julien Grenet (École d'économie de Paris), Pascal Bressoux (Université Grenoble-Alpes) et Peter Blatchford (University College London).

## 2.1 Objectifs de l'évaluation

Le protocole d'évaluation s'est appuyé sur les enseignements des études et recherches déjà menées en France et à l'international pour évaluer l'impact de la réduction d'une taille de classe et décrites à la section précédente. Ses objectifs étaient les suivants :

1. Évaluer l'impact à court terme ;
2. Évaluer l'impact à moyen terme (N+1 et N+2) ;
3. Décrire précisément les modalités de déploiement du dispositif ;
4. Recueillir des informations sur les pratiques d'enseignement ;
5. Observer l'engagement des élèves et les interactions professeur-élèves.

*Objectif 1. Évaluer l'impact à court terme.* Il s'agissait ici d'évaluer l'impact à court terme de la mesure de réduction de la taille des classes de CP en REP+, du point de vue de la progression des élèves de ces classes par rapport à d'autres élèves de CP. Pour ce faire, la DEPP a mis en place des évaluations standardisées en CP qui ont été passées par les élèves de l'échantillon. Cet échantillon est composé de 204 écoles de REP+, de 102 écoles de REP (ces écoles ont été choisies car elles présentaient des caractéristiques très proches de celles des écoles de REP+) et de 102 écoles publiques hors EP. Tous les élèves de CP de ces écoles sont concernés par les évaluations, soit un échantillon total de 15 000 élèves de CP. Les évaluations sur support numérique ont été conduites sur tablettes et comportaient quatre séquences de 20 minutes chacune : deux séquences de français, une séquence de mathématiques et une séquence composée d'items conatifs (estime de soi, bien-être, intérêt pour les apprentissages). Deux temps de mesure ont été organisés : le premier en septembre 2017 et le second en juin 2018.

*Objectif 2. Évaluer l'impact à moyen terme (N+1 et N+2).* La DEPP a également mis en place des évaluations standardisées en CE1 et en début de CM1 (l'évaluation initialement prévue en fin de CE2, en juin 2020, ayant été reportée en raison de la crise sanitaire). Ici deux échantillons d'élèves sont concernés. En effet, en plus de l'échantillon CP12 (décrit dans l'objectif 1), la DEPP a constitué un autre échantillon de comparaison, formé de tous les élèves de CE1 à la rentrée 2017/2018 des écoles de l'échantillon CP12, soit environ 15 000 élèves. Par construction, ces élèves de CE1 n'ont donc pas connu la réduction de la taille des classes. La DEPP a fait passer à ces élèves de CE1 des évaluations en novembre 2017, puis en juin 2018. Ces évaluations ont également été passées par les élèves de CP de CP12 lorsqu'ils sont arrivés en CE1, donc en novembre 2018 et en juin 2019. Cette cohorte a en outre passé des évaluations en début de CM1, en octobre 2020.

*Objectif 3. Décrire précisément les modalités de déploiement du dispositif.* La DEPP a interrogé *via* un questionnaire en ligne un échantillon de directeurs d'école sur l'organisation de l'école en décembre 2017. L'échantillon est composé de 1 222 écoles de REP+, 568 écoles de REP et 628 écoles hors EP. Il comprend l'ensemble des 408 écoles suivies dans le cadre de l'évaluation des apprentissages des élèves. En allant au-delà de l'échantillon CP12, il permet de disposer de suffisamment de descriptions pour rendre compte du déploiement de la mesure au niveau national.

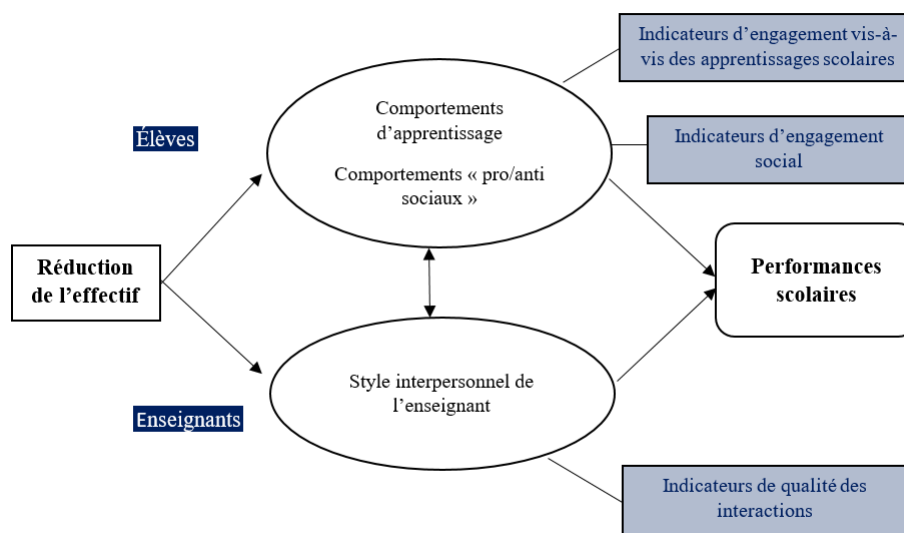
*Objectif 4. Recueillir des informations sur les pratiques d'enseignement.* Il s'agissait d'apprécier l'évolution des pratiques d'enseignement en classe dédoublée par le biais de questionnaires auto-rapportés conçus avec l'appui du Laboratoire de recherche sur les apprentissages en contexte (LaRAC) de l'Université Grenoble-Alpes, auprès d'un échantillon composé de 1 219 écoles de REP+, 568 écoles de REP et 624 écoles hors EP. Les enseignants interrogés dans ce cadre sont ceux des élèves de l'échantillon CP12 (soit leurs enseignants de CP en 2017/2018, puis leurs enseignants de CE1 en 2018/2019). La DEPP a également constitué un panel d'enseignants de



CP suivis de façon longitudinale (mai 2018, 2019, 2021 ) afin d'étudier d'éventuels effets d'ajustement dans les pratiques qui seraient différés dans le temps par rapport au calendrier de mise en œuvre de la politique. Afin de fournir une photographie suffisamment précise des pratiques au niveau national, le panel est composé de 4 600 enseignants de REP+, de 1 300 enseignants de REP et de 1 200 enseignants hors EP et inclut l'échantillon des enseignants de CP des écoles CP12 en 2017/2018.

*Objectif 5. Observer l'engagement des élèves et les interactions professeur-élèves.* L'objectif était de recueillir, grâce à un protocole d'observation *in situ* conçu par le LaRAC en lien avec Peter Blatchford (un chercheur anglais spécialisé sur la question de la taille des classes), des informations sur deux dimensions : l'engagement des élèves et les interactions entre les élèves et le professeur. Dans leur synthèse de 2003, Finn, Pannozzo & Achilles indiquent que l'effet de la réduction de l'effectif sur les performances scolaires des élèves transite à la fois par des modifications de comportements d'apprentissage et de comportements « sociaux » des élèves ainsi que par la modification du style interpersonnel des enseignants, donc de la façon dont les enseignants interagissent avec leurs élèves (Figure 1). Ainsi, pour comprendre comment la diminution de l'effectif peut agir sur les performances scolaires, il est nécessaire de se doter d'une méthode et d'indicateurs pouvant permettre d'identifier des modifications dans les comportements d'apprentissage des élèves, dans leurs comportements sociaux au sein de la classe et d'indicateurs permettant de détecter des fluctuations dans la façon dont les enseignants interagissent avec leurs élèves, que ce soit au niveau de la fréquence ou de la qualité des interactions. Dans ce but, des observations ont été menées en classe pour tenter de détecter des différences potentielles dans les comportements d'apprentissage des élèves et leurs comportements « sociaux », d'une part, et dans les pratiques d'interactions interpersonnelles de leur enseignant, d'autre part, selon qu'ils étaient dans des classes réduites ou non. Les observations de classe ont été répétées pendant trois ans auprès de 150 classes de CP de REP+, REP et hors EP, réparties dans 80 écoles appartenant à l'échantillon du volet « pratiques enseignantes » décrit à l'objectif 4.

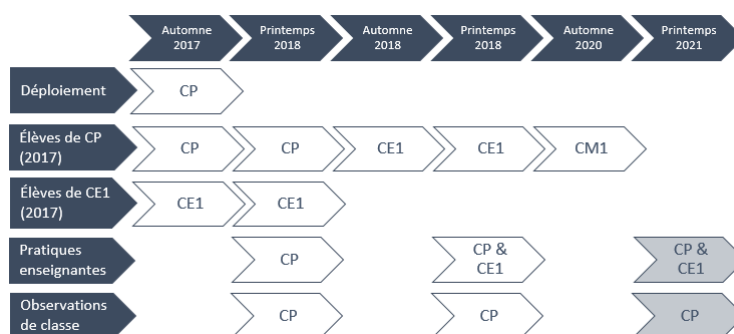
FIGURE 1 – Modèle de l'effet de la réduction de l'effectif des classes (à partir de Finn, Pannozzo & Achilles, 2003)



La Figure 2 récapitule l'ensemble des collectes de données prévues afin de réaliser les 5 objectifs d'évaluation susmentionnés. Le protocole initial prévoyait un suivi en CP, puis en CE1 et CE2, sur une période de trois années scolaires consécutives. Comme décrit précédemment, des

ajustements, par rapport aux collectes prévues au printemps 2020, sont intervenus du fait de la crise sanitaire.

FIGURE 2 – Cibles et calendrier des collectes de données de l'évaluation)



## 2.2 Méthodologie d'échantillonnage

Cette partie décrit les différentes étapes de la constitution des échantillons d'écoles pour le volet « élèves ».

### 2.2.1 Base de sondage

La base de sondage est issue d'une extraction d'APAE. Elle couvre au départ l'ensemble des écoles publiques en France métropolitaine et dans les DOM, ce qui représente 45 918 écoles. On passe à 45 736 en enlevant Mayotte, à 27 687 en ne retenant que les écoles avec au moins un élève de CP et à 20 003 en ne retenant que les écoles avec au moins 13 élèves de CP.

### 2.2.2 Contraintes du tirage

- La taille et la structure de l'échantillon ont été définies par plusieurs contraintes :
- La passation des évaluations sous forme numérique impose de fortes contraintes logistiques pour tenir compte de la répartition des mallettes de tablettes en académie. On estime à 4 environ le nombre d'écoles que l'on pourra évaluer avec une mallette en deux semaines de passation. La taille maximale de l'échantillon des écoles pour l'évaluation des élèves est 408 écoles.
  - L'objectif principal de cette opération était d'évaluer la réduction de la taille des classes en REP+, en tenant compte de l'existence du dispositif antérieur « Plus de maîtres que de classes » (PMQC) également mis en œuvre en REP+. On a donc décidé de constituer une strate de REP+ avec la moitié des écoles de l'échantillon, soit 204 écoles. On souhaitait disposer par ailleurs d'un groupe témoin le plus comparable possible, tiré par exemple parmi les REP les plus proches des REP+, dont on a fixé la taille à un quart de l'échantillon, soit 102. En effet, on a réservé un quart de l'échantillon pour le reste de la population.
  - À la rentrée 2017, une partie des écoles devaient commencer le dédoublement des classes, d'autres devaient maintenir le dispositif PMQC, avec des situations intermédiaires entre ces deux options. L'entrée dans le dispositif CP12 pouvait être sans doute freinée par des questions de locaux. Pour essayer de contrôler un peu le partage non aléatoire entre les deux dispositifs, on a tiré parti de l'évaluation déjà menée pour PMQC par la DEPP, dans 4 départements. Les écoles de cette évaluation dans les départements du Nord et du Rhône ont été choisis en priorité.

Pour respecter les deux premières contraintes, on a estimé le nombre d'écoles à tirer dans chaque strate pour chaque académie en appliquant les paramètres  $\frac{1}{2}$ ,  $\frac{1}{4}$  et  $\frac{1}{4}$  aux nombres d'écoles évaluables de chaque académie. Le nombre ainsi obtenu est généralement inférieur au nombre total d'écoles appartenant à la strate dans l'académie, à deux exceptions près :

- À Rennes, on estime à 10 le nombre d'écoles REP+ à interroger, alors qu'il n'y en a effectivement que 2.
- En Guyane, il n'y a plus aucune école en REP alors que le plan théorique demande d'en évaluer 3.

On a résolu ces problèmes de la façon suivante :

- À Rennes, on n'a interrogé que 2 écoles en REP+ ; on a reporté les 8 écoles évaluables équitablement dans les deux autres strates.
- En Guyane, on a interrogé 8 écoles en REP+ (au lieu de 6 dans le plan initial) et 4 écoles hors EP (soit toutes les écoles hors EP de l'académie).
- À Lille et à Lyon, où s'est faite l'évaluation de PMQC, on a gonflé l'échantillon de 3 REP+ dans chaque académie et on a passé dans les deux autres strates à 8 et 5 à Lille et 3 et 2 à Lyon.

### 2.2.3 Construction d'un groupe comparable au groupe REP+

Pour constituer le groupe témoin, nous avons mis en œuvre une procédure d'appariement global. En effet, une procédure d'appariement individuel risquait de donner des résultats assez médiocres, car pour les écoles de REP+ les plus défavorisées, il est difficile de trouver un équivalent dans le reste de la population.

Dans tous les cas, il faut d'abord établir un critère pour définir cet échantillon témoin. Nous avons pris trois caractéristiques intéressantes des écoles : la proportion d'élèves de PCS défavorisés en CM2 (issue des fichiers du second degré), le taux de retard en CM2 et le revenu médian de l'IRIS où se trouve l'école (ces variables étaient manquantes pour certaines écoles : dans ce cas, des imputations ont été réalisées, notamment à l'aide de celles de ces 3 variables qui sont renseignées). Nous avons construit un indicateur global avec ces trois informations en effectuant la régression logistique de l'appartenance au réseau REP+. Cela donne pour chaque école une « propension » estimée d'appartenir à ce réseau, combinaison linéaire des trois informations. Cette propension s'avère un bon prédicteur de l'appartenance effective au réseau. En effet, nous avons découpé la population en 40 tranches égales par ordre croissant de propension. Dans les 27 premières tranches, la proportion d'écoles en REP+ est inférieure à 1 % ; dans les 6 dernières tranches, on dépasse 10 %, avec notamment des valeurs de 46 % et 69 % dans les deux dernières tranches. Cette propension peut donc nous servir à construire un groupe témoin.

On souhaitait tirer en priorité des REP, mais il n'y en avait pas assez dans la dernière tranche, même si l'on ajoutait les écoles comparables hors EP (155 contre 345 en REP+). On a décidé de tirer toutes ces écoles, même hors EP et on a calculé le nombre d'écoles à tirer en REP dans les autres tranches en calant la répartition selon la propension sur celle en REP+. On obtient ainsi un groupe de 550 écoles comparables globalement aux REP+ suivant les critères retenus.

### 2.2.4 Tirage de l'échantillon

On dispose maintenant d'une partition de la population dans les trois strates d'intérêt : REP+, groupe comparable (REP, qualifié de « Proche REP+ ») et reste de la population. On a précédemment déterminé combien d'écoles par académie pouvaient être tirées dans ces strates. Il n'y a plus alors qu'à effectuer le tirage, en ajoutant des critères d'équilibrage, qui seront ici les valeurs moyennes des trois variables utilisées pour créer le groupe comparable (pourcentage

de PCS défavorisées, revenu médian de l'IRIS, taux de retard en CM2) et la distribution dans les quarante tranches de propension dans les groupes REP+ et REP comparables. Dans les académies du Nord et de Lille, les écoles sélectionnées en REP+ ont été remplacées par des écoles ayant déjà participé à l'évaluation PMQC.

### 3 Mise en œuvre de la politique évaluée

Une évaluation d'impact de politique publique ne peut faire l'économie d'une description précise des modalités de mise en œuvre de la politique afin de pouvoir mettre en regard les effets de la politique avec les actions effectivement déployées. Cette section s'intéresse aux données administratives et aux données d'enquête mobilisées par la DEPP à cette fin. Ainsi, les données exhaustives présentes dans les systèmes d'information du Ministère permettent de décrire l'ampleur effective de la réduction de la taille de classe et son incidence sur la structure des classes. Les données d'enquête collectées auprès d'échantillons de directeurs d'école et d'enseignants mettent en lumière d'autres aspects clés du déploiement de la politique, comme le partage des locaux, l'accompagnement réalisé par l'IEN (inspecteur de l'Éducation nationale) de circonscription ou la formation des enseignants.

À titre d'illustration, les éléments fournis ci-après portent sur les conditions d'enseignement et d'apprentissage observées en 2017/2018, quand les élèves de l'échantillon évalué étaient en classe de CP. Le lecteur intéressé par la description de la mise en œuvre de la réforme en 2018/2019 (quand ces mêmes élèves étaient scolarisés en CE1) pourra se référer à l'étude complète de la DEPP (Andreu *et al.* 2021).

La mise en place progressive de la mesure dite de « dédoublement » de la taille des classes en éducation prioritaire prévoyait une mise en œuvre pour le niveau CP en REP+ dès la rentrée de septembre 2017. Les données du constat ont été utilisées par la DEPP pour quantifier l'ampleur de la réduction de la taille de classe au niveau national. En REP+, la taille moyenne des classes de CP est passée de 21,6 élèves à la rentrée 2016 à 13,1 en 2017 (Tableau 1). À titre de comparaison, en REP, en septembre 2017, les classes de CP comptent en moyenne 20,7 élèves. La mise en œuvre de la mesure de réduction des tailles de classes de CP est également perceptible *via* l'évolution de la proportion de classes de CP multiniveau. En septembre 2017, la proportion de ces classes en REP+ chute à 5 %, contre 25 % l'année précédente et 28 % en REP la même année.

TABLEAU 1 – Évolution de la taille moyenne des classes de CP

	REP+	REP	Hors EP
Rentrée 2016	21,6	21,6	22,6
Rentrée 2017	13,1	20,7	22,3
Rentrée 2018	12,7	12,8	22,3

Champ : France métropolitaine + DROM, écoles publiques, hors élèves ULIS.  
Source : MENJS-DEPP, bases Diapre 2016 à 2018.

Une enquête *ad hoc* menée auprès des directeurs d'école en décembre 2017 a également permis de mettre en évidence le pilotage local de la réforme. En 2017, plus de huit directeurs sur dix considèrent que la mise en place du dispositif « CP dédoublés » a fait l'objet d'une impulsion donnée par l'IEN, en lien avec la DSDEN (direction des services départementaux de l'Éducation nationale). Le rôle de l'IEN de circonscription apparaît notamment dans la sélection des enseignants affectés aux classes dédoublées en CP : environ la moitié des directeurs d'école

de REP+ rapporte avoir bénéficié d'une implication de l'inspecteur pour l'attribution des classes « dédoublées » à tout ou partie des enseignants ayant effectivement pris en charge une classe à effectif réduit dans l'école. Les résultats de cette enquête permettent également de décrire les critères ayant prévalu à la rentrée 2017 pour l'attribution des classes de CP aux enseignants selon l'appartenance ou non à un réseau d'éducation prioritaire (Tableau 2). Ainsi, pour près de huit directeurs de REP+ sur dix, l'expérience en classe de CP a été déterminante dans l'attribution des classes de ce niveau aux enseignants, contre environ six directeurs sur dix en REP. L'expérience en éducation prioritaire ainsi que dans le dispositif « Plus de maîtres que de classes » ont également été des critères plus prépondérants dans ce choix en REP+ qu'en REP.

TABLEAU 2 – Proportion de directeurs d'école indiquant que les critères suivants ont prévalu pour l'attribution des classes de CP aux enseignants à la rentrée 2017

	REP		
	REP+	(Proche REP+)	Hors EP
Expérience en classe de CP	79	62	59
Expérience en contexte d'éducation prioritaire	60	44	3
Expérience dans « Plus de maîtres que de classes »	46	22	2
Ancienneté générale	27	32	25
Volontariat	85	84	82
Quotité de service	23	20	14
Nécessité de pallier une absence	11	10	5

Note : La moyenne REP est calculée à partir des réponses des directeurs d'école de REP sélectionnées dans l'échantillon d'enquête car elles présentaient des caractéristiques similaires à celles des écoles de REP+ (il s'agit donc d'un sous-ensemble particulier, non représentatif des écoles REP au niveau national). C'est le cas de tous les résultats tirés des données d'enquête.

Source : MENJS-DEPP, enquête auprès des directeurs d'école (décembre 2017).

En 2017/2018, les enseignants des CP dédoublés en REP+ sont les plus nombreux à déclarer avoir bénéficié d'un accompagnement soutenu de la circonscription pour la prise en main de la classe de CP (48 % d'entre eux vs 12 % en REP et 5 % hors EP).

La mise en œuvre de la politique s'est accompagnée d'un surcroît de formation des enseignants affectés aux classes de CP dédoublés en 2017/2018 (Tableau 3). Lors de l'enquête administrée au

TABLEAU 3 – Proportion d'enseignants ayant bénéficié d'une formation sur le niveau CP en 2017/2018

	REP		
	REP+	(Proche REP+)	Hors EP
En circonscription	93	66	59
Au niveau de la DSDEN	40	17	11
Au niveau de l'académie	30	9	5
Sur l'apprentissage de la lecture	85	49	32
Sur l'enseignement des mathématiques	54	36	36

Source : MENJS-DEPP, enquête auprès des enseignants de CE1 (mai-juin 2019).

printemps 2018, les enseignants des CP dédoublés en REP+ sont plus nombreux que les autres enseignants de CP interrogés dans le secteur public (exerçant en REP ou hors EP) à déclarer avoir bénéficié au cours de l'année d'une formation d'au moins 6 heures sur le niveau CP. Des écarts

étaient déjà observés sur la formation en 2016-2017, mais ceux-ci étaient de moindre ampleur. En REP+, plus de neuf enseignants de CP sur dix rapportent avoir bénéficié d'une formation sur le niveau CP organisée en circonscription au cours de l'année. Ils sont également 40 % à déclarer avoir bénéficié d'une formation au niveau départemental et 30 % au niveau académique. Ils sont 1,7 fois plus nombreux que leurs collègues de REP à avoir bénéficié d'une formation sur l'apprentissage de la lecture (85 % *vs* 49 %). L'écart est également important s'agissant de la participation à une formation portant sur l'enseignement des mathématiques (54 % de bénéficiaires d'une formation en REP+ *vs* 36 % en REP).

## 4 Mesure de la progression des élèves

Les données récoltées par la DEPP permettent d'étudier l'effet de la réduction de la taille des classes sur la réussite des élèves. Pour cela, les résultats des élèves en REP+, qui ont bénéficié du dispositif dès la rentrée 2017, sont comparés à ceux d'élèves qui n'en ont pas bénéficié, mais qui ont des caractéristiques semblables. Ces élèves se trouvent dans des écoles, dites « Proches REP+ » qui ont été sélectionnées par la DEPP afin de constituer un ensemble dont les caractéristiques sociales et scolaires sont en moyenne proches de celles des écoles de REP+ bénéficiaires de la réduction de la taille des classes.

Le « traitement » dont on va ici évaluer l'effet sur les compétences des élèves est le fait d'appartenir à une école de REP+. En effet, la taille effective des classes n'a pas été prise en compte, car même en se restreignant à des écoles d'éducation prioritaire, il y a un risque d'« endogénéité » : quelques écoles de REP+ n'ont pas bénéficié de la réduction de la taille des classes ; certaines écoles du groupe témoin, en REP, ont des effectifs faibles ; il est probable que dans les deux cas, ces écoles aient des caractéristiques particulières, ce qui va biaiser la comparaison. Cette étude retient donc le principe d'une comparaison globale : les élèves en REP+ ont, en moyenne, bénéficié d'une réduction sensible de la taille des classes ; ils seront comparés à des élèves leur ressemblant sur des critères sociaux et scolaires, qui, en moyenne, n'ont pas bénéficié de cette mesure. Par la suite, il sera possible de préciser l'effet de la mesure sur les compétences des élèves, par l'intensité du traitement (la différence de tailles des classes entre les groupes).

La mesure de dédoublement visait une réduction de la taille des classes dans les établissements de REP+ en CP dès 2017. En effet, entre les années scolaires 2016/2017 et 2017/2018, on observe une baisse de 8 élèves en moyenne pour les établissements REP+ de l'échantillon pour atteindre un peu moins de 13 élèves (Tableau 4). Une baisse équivalente est observée pour les classes de CE1 en 2018/2019 (un peu moins de 14 élèves par classe contre un peu plus de 20 l'année précédente). De plus, on a une différence de 7 élèves entre les REP+ et les Proches REP+ en CP et en CE1.

On va ici décrire l'évolution des compétences du début du CP à la fin du CE1, des élèves entrés en CP en 2017, en s'attachant à comparer les élèves de REP+, ayant bénéficié de la réduction de la taille des classes, et les élèves des écoles Proches REP+ qui n'ont pas bénéficié. Soit 10 226 élèves au total en français et 7 801 élèves en mathématiques pour lesquels on dispose d'informations à tous les temps de mesure. Il est à noter que l'attrition entre les différents temps d'observation était plus importante en mathématiques qu'en français, car cette épreuve était généralement proposée en dernier lors de la passation, ce qui a pu, en cas de difficultés, provoquer plus souvent son abandon. Cependant, l'attrition était proportionnelle entre les groupes. De plus, pour la majorité des temps d'observation, les réussites sur le champ restreint ne montrent pas de différence significative avec le périmètre considérant la totalité des élèves évalués.

L'attention portera notamment sur la progression des scores de réussite moyens aux épreuves

TABLEAU 4 – Taille des classes de CP et de CE1 de l'échantillon par groupe de 2016 à 2018

	CP			CE1		
	2016/17	2017/18	2018/19	2016/17	2017/18	2018/19
Hors EP	23,3	23	22,5	24	24,4	23,7
REP (Proche REP+)	21,7	20	12,6	22,4	21,9	21
REP+	21,2	12,7	12,6	21,8	20,3	13,7

Lecture : Durant l'année scolaire 2017/2018, la taille moyenne des classes de l'échantillon avec au moins un élève de CP en REP+ est de 12,7 élèves.

Source : MENJS-DEPP, enquête auprès des élèves.

de mathématiques et de français. Des modélisations économétriques seront effectuées, pour contrôler l'influence d'autres facteurs que la taille des classes, liés à la réussite scolaire.

## 4.1 Description des données disponibles

### 4.1.1 Modalités d'évaluation

Les élèves ont été évalués grâce à une application numérique développée sur tablette. À chaque temps d'évaluation, la DEPP a recueilli des informations sur les élèves et notamment les caractéristiques nécessaires aux analyses : sexe, date de naissance, classe et école fréquentées.

### 4.1.2 Domaines évalués en français

Lors de chaque temps de mesure (début de CP, fin de CP, début de CE1 et fin de CE1), ils ont passé 2 épreuves de 20 minutes effectives de travail chacune, proposées en deux séances lors de deux demi-journées distinctes. Les items évaluent différents éléments du programme d'enseignement du cycle des apprentissages fondamentaux (cycle 2, BO n°11 du 26 novembre 2015) associées aux compétences du socle commun de connaissances, de compétences et de culture.

La lecture est une activité complexe dont la maîtrise implique la mise en œuvre simultanée de deux principales activités (Gough & Turner, 1986) : l'identification des mots écrits et la compréhension de la signification de ces mots. Aussi, plusieurs compétences sous-tendent l'efficacité de la mise en place de ces deux activités. D'une part, la connaissance de lettres, les habiletés phonologiques sont liées au développement des processus d'identification et de production de mots écrits (e.g., Puolakanaho *et al.*, 2007). D'autre part, le niveau de vocabulaire et la compréhension orale de textes sont reliés à la compréhension en lecture (e.g., Cain & Oakill, 2007). Précisément, ces compétences sont reconnues comme de solides prédicteurs de la réussite lors de l'apprentissage de la lecture. Par conséquent, un faible niveau dans ces compétences prédictives suggère des risques éventuels de difficultés ultérieures dans le développement des capacités de lecture.

Ainsi, plusieurs champs sont investigués, les habiletés phonologiques, la connaissance des lettres, la compréhension orale de récit et de vocabulaire, la lecture de mots et de pseudo-mots. À chaque temps de mesure, des items du temps précédent sont repris à l'identique.

### 4.1.3 Domaines évalués en mathématiques

À chaque temps de mesure, les élèves se sont vu proposer 1 épreuve de 20 minutes effectives de travail, proposée en une séance. Les domaines évalués, ici aussi associés aux compétences du socle commun de connaissances, de compétences et de culture, sont les suivants : nombres et calculs, grandeurs et mesure, espace et géométrie. Comme en français, à chaque temps de mesure, des items du temps précédent sont repris à l'identique.

## 4.2 Progressions des élèves entre le début de CP et la fin de CE1

### 4.2.1 Statistiques descriptives

Pour faciliter les comparaisons, sur des ensembles d'items variant d'un temps d'observation à l'autre, nous avons calculé des scores standardisés. Le score obtenu a été fixé par construction à une moyenne de 0 et un écart-type de 1 aux différents temps de passation pour le groupe Proche REP+. Les valeurs s'interprètent alors comme l'écart (en % d'écart-type) par rapport au groupe témoin. Une valeur négative signifie que les REP+ ont des résultats inférieurs aux Proches REP+ ; une valeur positive indique au contraire que leurs résultats sont meilleurs.

En français, les écarts de réussite entre les REP+ et les Proches REP+ en fin de CE1 sont du même ordre en fin de CE1 et en début de CP (environ 6 % d'écart-type). En revanche, la réduction de la taille des classes semble avoir un effet positif sur les résultats en mathématiques des élèves de REP+, avec une différence de performance de l'ordre de 8,0 % d'écart-type par rapport aux Proches REP+ en fin de CE1, là où il n'y avait aucun écart en début de CP (Tableau 5).

TABLEAU 5 – Scores standardisés entre le début du CP et la fin du CE1 des REP+ par rapport aux Proches REP+

	Français		Mathématiques	
	Moyenne	Écart-type	Moyenne	Écart-type
Début de CP	-0,068***	-1,02	0	-1,02
Fin de CP	-0,054**	-1,04	0,048*	-1,01
Début de CE1	-0,029	-1,04	0,050 *	-1,01
Fin de CE1	-0,062**	-1,02	0,080***	-0,98

Note : Significativité de la différence par rapport au groupe Proche REP+ : p-value < 1 % : \*\*\*, p-value < 5 % : \*\*, p-value < 10 % : \*. (0,00) correspond à l'erreur standard associée au coefficient. Lecture : Le score standardisé sur l'ensemble des items de français des REP+ en début de CP est inférieur significativement à celui des Proches REP+ de 6,8 % d'écart-type. Source : MENJS-DEPP, enquête auprès des élèves.

### 4.2.2 Estimation économétrique de l'effet de la réduction de la taille des classes

Ces statistiques descriptives donnent une première idée de l'impact de la réduction de taille des classes en REP+, mais il est nécessaire de procéder à des modélisations statistiques pour contrôler l'influence de certains facteurs liés à la progression des élèves, et sur lesquels REP+ et Proches REP+ peuvent se distinguer. Les données disponibles ne sont pas très riches, mais il a été possible de prendre en compte le sexe et le retard scolaire de l'élève à l'entrée en CP, ainsi



que le département <sup>6</sup> dans lequel est située l'école. De plus, les modèles sont déclinés par discipline.

L'enjeu de la modélisation est de comparer des populations proches. C'est pourquoi les élèves scolarisés dans les écoles hors éducation prioritaire sont écartés des modélisations.

Les modèles régressent la différence entre le score en fin et le score en début de période étudiée par le groupe d'appartenance de l'élève et les variables de contrôle <sup>7</sup>. De façon générale, dans ces modèles, on observe une progression <sup>8</sup> moins importante des garçons, que ce soit en français ou en mathématiques <sup>9</sup>; les élèves en retard à l'entrée en CP progressent moins que les autres; il existe aussi des évolutions assez différentes d'un département à l'autre (résultats non présentés en détail ici et qui pourraient sans doute être étudiés de façon plus précise avec les évaluations exhaustives en début de CP et début de CE1).

Ces modèles donnent une estimation plus favorable de l'effet de la réduction des tailles de classe que les statistiques descriptives : en mathématiques, les élèves en REP+ ont une progression de 14 % d'écart-type supérieure à celle des Proches REP+, alors que les statistiques indiquaient une amélioration de 8 % en faveur des REP+ (Tableau 6). En français, un effet de 9 % d'écart-type est observé, significatif au seuil de 10 %, alors qu'aucune différence n'apparaissait dans les statistiques descriptives. C'est la prise en compte d'une répartition différenciée des REP+ et des Proches REP+ par département qui explique surtout cette divergence entre statistiques descriptives et modélisation.

En mathématiques, la progression de 14 % correspond à un effet équivalent à 38 % de l'écart observé en début de CP entre le groupe REP+ et le groupe Hors EP (37 % d'écart-type). En français, la progression de 9 % correspond à 16 % de l'écart observé en début de CP entre le groupe REP+ et le groupe Hors EP (57 % d'écart-type).

Rapporté à un écart de 7 élèves dans la taille des classes des deux groupes comparés, cela revient à un effet de 1,3 % d'écart-type par élève en moins en français et 2 % en mathématiques. Par rapport aux résultats trouvés antérieurement, l'effet paraît un peu faible en français, mais conforme aux attentes en mathématiques. En mathématiques, cet effet paraît plus fort pour les élèves les plus en difficultés : leur part passe de 21,4 % au début du CP à 15,9 % en fin de CE1 (le niveau ayant été fixé à 20 % dans les écoles non bénéficiaires de la politique, comparables aux REP+), alors que l'évolution est très faible pour la part d'élèves très performants.

Les analyses complémentaires menées par la DEPP (Andreu *et al.*, 2021) montrent que l'impact positif de la réduction de la taille des classes en REP+ est surtout visible en CP (+14 %

---

6. La répartition des écoles en réseau d'éducation prioritaire n'est pas strictement comparable sur l'ensemble du territoire. De façon à prendre en compte cet effet de structure, le département dans lequel se situent les écoles de l'échantillon a été intégré aux analyses. Tous les modèles statistiques du document prennent en compte le département, mais les coefficients ne sont pas affichés dans les tableaux de résultats pour des raisons de présentation.

7. Pour tenir compte de la structure des données (regroupement des élèves en classes et des classes en écoles), ce sont des modèles multiniveaux, à effets aléatoires qui ont été utilisés. Le lecteur intéressé trouvera dans Andreu *et al.* (2021) en annexe 8 une discussion des choix méthodologiques en matière de modélisation économétrique et en annexe 9 les résultats d'une autre méthode, donnant des résultats très proches.

8. Comme il s'agit de scores standardisés, dont la moyenne est fixée à 0 pour les Proches REP+, une valeur négative de la progression (de la différence entre son score en t et son score en t-1) ne signifie pas que l'élève régresse, mais qu'il progresse moins que l'élève moyen de Proches REP+ et qu'il perd des places dans son classement selon le score dans la discipline.

9. Le creusement des écarts entre garçons et filles en faveur des secondes en français est connu. En revanche, le fait que les garçons progresseraient moins que les filles en mathématiques n'est généralement pas observé et devra être analysé plus en détail.

TABLEAU 6 – Ecart de réussite en français et en mathématiques entre le début de CP 2017 et la fin de CE1 2018 en fonction du groupe, du sexe, du retard scolaire et du département

	Français		Mathématiques	
	Modèle 1	Modèle 2	Modèle 1	Modèle 2
Constante	-0,01 (-0,04)	-0,04 (-0,23)	0 (-0,04)	0,2 (-0,21)
Groupe REP+ (réf. Proches REP+)	-0,01 (-0,05)	0,09* (-0,05)	0,08* (-0,04)	0,14*** (-0,05)
Sexe (réf. Filles)		- 0,25*** (-0,02)		- 0,15*** (-0,02)
Retard (réf. « À l'heure »)		- 0,28*** (-0,07)		- 0,50*** (-0,09)
Département		Oui		Oui
Effets aleatoires				
Variance inter élève	0,88	0,86	0,85	0,84
Variance inter classe CP	0,02	0,02	0,02	0,01
Variance inter classe CE1	0,02	0,02	0,04	0,04
Variance inter école	0,11	0,07	0,05	0,04
Nombre d'observations :				
Elèves	7 762		5 830	
Classes de CP	851		838	
Classes de CE1	816		794	
Ecoles	288		284	

Note : Significativité de la différence par rapport au groupe Proche REP+ : p-value < 1 % : \*\*\*, p-value < 5 % : \*\*, p-value < 10 % : \*. (0,00) correspond à l'erreur standard associée au coefficient.

Lecture : En français, à caractéristiques comparables présentes dans le modèle, les élèves en REP+ progressent davantage que les élèves en Proches REP+, entre le début de CP et la fin de CE1, puisque le coefficient estimé est positif (0,09) et significatif.

Source : MENJS-DEPP, enquête auprès des élèves.

d'écart-type en français et +12 % d'écart-type en mathématiques). En CE1, il ne semble pas y avoir d'effet supplémentaire (le différentiel de progressions entre REP+ et proches REP+ est faible et non significatif, tant en français qu'en mathématiques).

## 5 Mesure de l'évolution des pratiques et du vécu des enseignants

La littérature empirique sur la réduction de la taille de classe s'intéresse principalement à la mesure des effets sur la réussite scolaire des élèves (Monso, 2014 ; Bouguen, Grenet & Gurgand, 2017). Afin de documenter les mécanismes susceptibles de lier la réduction d'effectifs en classe à la performance des élèves, la DEPP a administré des enquêtes auprès d'enseignants de CP et de CE1 en REP+, en REP et dans le secteur public hors éducation prioritaire (EP) aux printemps 2018, 2019 et 2021. Ces enquêtes devaient notamment permettre à la fois de :

- décrire les pratiques et le ressenti des enseignants des élèves des classes de CP (en 2017/2018), puis de CE1 (en 2018/2019) bénéficiaires de la mesure de réduction de la taille en REP+ en les comparant aux déclarations d'enseignants présentant des caractéristiques similaires et exerçant dans des contextes proches, mais non bénéficiaires de la mesure évaluée<sup>10</sup> ;
- et d'étudier l'évolution des pratiques et du ressenti des enseignants de CP en REP+ entre la première et les années ultérieures de mise en œuvre de la mesure *via* la constitution d'un panel de 4 600 enseignants de REP+, de 1 300 enseignants de REP et de 1 200 enseignants hors EP afin de rendre compte d'éventuels effets d'ajustement des pratiques qui pourraient être différés dans le temps (par rapport à la mise en place de la mesure évaluée).

Les enseignants sollicités pour ces enquêtes exerçaient soit dans les écoles concernées par l'évaluation des élèves (Section 4), soit dans d'autres écoles situées en REP+, REP ou hors EP. Ainsi, les données permettaient, d'une part, de décrire les conditions d'enseignement et d'apprentissage dans les classes des écoles de l'échantillon des élèves et, d'autre part, d'aller au-delà de cet échantillon restreint afin de fournir une photographie plus précise des conditions d'enseignement et d'apprentissage au niveau national après la mise en place de la réforme. Au total, l'échantillon d'écoles concernées par les enquêtes auprès des enseignants de CP et de CE1 comptait 1 219 écoles de REP+, 568 écoles de REP et 624 écoles hors EP.

Le questionnaire a été conçu par les chercheurs Pascal Bressoux et Laurent Lima (LaRAC, Université Grenoble-Alpes) associés à l'évaluation afin de recueillir des informations sur les pratiques en classe des enseignants et leur expérience d'enseignant. D'une durée de 45 minutes environ, il était auto-administré en ligne *via* le logiciel d'enquête LimeSurvey.

### 5.1 Domaines interrogés

L'enquête incluait des questions sur les caractéristiques individuelles des enseignants, comme leur niveau de diplôme, leur statut ou leur ancienneté (dans le métier et dans la prise en charge d'élèves de CP ou CE1), et leur contexte d'exercice, comme la taille et la composition scolaire de la classe prise en charge (mononiveau ou multiniveau) ou le partage de la salle de classe avec

---

10. Comme pour l'évaluation des effets de la mesure sur les progressions des élèves, les écoles du groupe de comparaison (« REP (proche REP+) ») ont été sélectionnées par la DEPP afin de constituer un ensemble dont les caractéristiques sont en moyenne proches de celles des écoles de REP+ bénéficiaires de la réduction de la taille des classes.

un autre collègue.

Plusieurs questions cherchaient à rendre compte des modalités de mise en place de la réforme au-delà de la seule réduction de taille de classe. Ainsi, le questionnaire interrogeait les enseignants sur leur accès à des activités de formation continue en leur demandant s'ils avaient bénéficié au cours de l'année scolaire écoulée (et de l'année antérieure) d'une formation d'au moins 6 heures sur le niveau enseigné (CP ou CE1) organisée en circonscription, au niveau départemental ou au niveau académique. Ils étaient également questionnés à propos des contenus de la formation éventuellement suivie (enseignement du français, des mathématiques et/ou d'un autre contenu). Enfin, une question portait sur l'intensité de l'accompagnement de la circonscription pour la prise en main de la classe en début d'année scolaire (visite, formation, etc.), puisque le questionnaire adressé aux directeurs d'école en décembre 2017 avait mis en lumière le rôle d'impulsion qu'avaient pu jouer les inspecteurs de circonscription.

Le regard porté par l'enseignant sur sa classe est susceptible d'influencer la relation enseignant-élève ainsi que ses pratiques pédagogiques (Bressoux, 2012). Par ailleurs, cette perception peut elle-même être influencée par le contexte scolaire (comme une classe de taille réduite). Aussi, le questionnaire soumis aux enseignants intégrait cette dimension en interrogeant les enseignants à propos du caractère globalement homogène ou hétérogène du niveau scolaire du groupe classe, du nombre d'élèves présentant des difficultés scolaires et/ou des difficultés de comportement. Selon la théorie sociocognitive issue des travaux du psychologue américain Albert Bandura (2010), le regard que l'enseignant porte sur lui-même (son sentiment d'efficacité personnelle, c'est-à-dire sa confiance en ses aptitudes à enseigner) est aussi un potentiel déterminant des pratiques enseignantes et de la perception qu'ont les élèves de leurs propres capacités. Trois domaines d'efficacité personnelle étaient couverts dans le questionnaire : la capacité à faire progresser tous les élèves, la capacité à gérer les comportements perturbateurs en classe et la capacité à mettre en place une différenciation pédagogique.

Les enseignants ont été interrogés à propos de leur ressenti s'agissant du fonctionnement de classe. Les questions mobilisées pour apprécier le climat de classe étaient tirées de l'enquête internationale TALIS (*Teaching and Learning International Survey*) de l'OCDE. On leur demandait, par exemple, à quelle fréquence les situations suivantes se produisaient dans leur classe : « *Il y a beaucoup de bruit qui perturbe la classe* » ou « *Au début de la séance, je dois attendre assez longtemps avant que les élèves se calment* ». La gestion de classe étant un enjeu majeur pour les enseignants français d'après les résultats de TALIS 2018 (Charpentier, Embarek, Raffaëli & Solnon, 2019), il était pertinent d'investiguer de potentiels effets de la mesure de réduction de la taille de classe sur le bien-être des enseignants.

Le questionnaire cherchait à identifier les convictions des enseignants en matière d'enseignement et d'apprentissage. Par exemple, en REP+, comme en REP, la quasi-totalité des enseignants de CP sont « d'accord » ou « tout à fait d'accord » avec l'affirmation selon laquelle « *C'est lorsque les élèves résolvent eux-mêmes des problèmes qu'ils apprennent le mieux* », ce qui renvoie à une conception de l'enseignement fondée plutôt sur une pédagogie de la découverte (qui est une méthode inductive d'enseignement s'appuyant sur les essais, les erreurs et le tâtonnement de l'élève pour construire les apprentissages). Si les convictions des enseignants informent sur l'adhésion des enseignants à certains modèles pédagogiques (notamment enseignement explicite *vs* enseignement par la découverte), elles ne sont pas nécessairement corrélées avec les pratiques mises en œuvre en classe. L'information ainsi recueillie permet peut-être de tester la comparabilité des groupes d'enseignants utilisés dans le cadre de l'évaluation, si elle s'apparente à une caractéristique individuelle non influencée par la politique mise en place. Dans la même veine, une question portait sur la conception que se font les enseignants de l'intelligence, qui peut être per-

gue comme une capacité relativement fixe et stable ou au contraire comme une capacité malléable.

Des questions relatives aux pratiques générales d'enseignement permettaient de décrire le style d'enseignement en classe. Les travaux en psychologie de l'éducation, comme ceux de l'Américaine Carol Dweck (2006), ont mis en évidence l'effet de certaines pratiques d'enseignement soutenant et orientant la motivation des élèves et de nature à agir sur les processus d'apprentissage en contexte scolaire. Aussi, le questionnaire incluait un ensemble de questions permettant de décrire le « climat motivationnel » (selon le terme usité dans la littérature) instauré par l'enseignant, c'est-à-dire les buts d'accomplissement favorisés par ses pratiques. Les réponses permettent de distinguer les interventions des enseignants favorisant les buts de maîtrise (fondés sur les expériences d'apprentissage, de maîtrise et de progrès personnel des élèves) de celles favorisant les buts de performance (fondés sur la comparaison et la compétition entre élèves). Des questions portaient sur la fréquence des démarches d'enseignement qui recherchent la stimulation cognitive des élèves, avec l'intégration dans les pratiques de tâches stimulantes, nécessitant un raisonnement élaboré, stimulant la curiosité des élèves ou posant des questions amenant les élèves à réfléchir en profondeur. Le questionnaire couvrait également la place de l'enseignement explicite (c'est-à-dire un mode d'enseignement caractérisé notamment par une démarche directe, structurée et fortement guidée par l'enseignant) et des pédagogies actives. Les questions étaient tirées de plusieurs échelles (dont celle développée par Bressoux, Lima & Nurra à partir notamment des travaux de Kyriakides, Christoforou & Charalambous (2013), celle développée par Nurra, Lepareur & Cross et celles utilisées dans l'enquête TALIS). Les enseignants étaient interrogés à propos de la fréquence de leur gestes professionnels en matière d'étayage, comme par exemple montrer aux élèves leurs progrès tout au long d'une activité, et de régulation de l'activité des élèves, comme par exemple donner des indices aux élèves au bout d'un certain temps ou simplifier la consigne quand les élèves sont en difficulté. D'autres questions portaient sur la fréquence des pratiques relevant de la différenciation pédagogique (c'est-à-dire l'adaptation de l'enseignement aux besoins des élèves), sur les modalités de mise en place du travail des élèves (travail individuel ou travail en groupe), l'utilisation de matériel numérique en classe et le suivi du travail personnel à la maison. La première année, les questions spécifiques aux contenus enseignés portaient uniquement sur l'enseignement du français. L'année d'après, un volet relatif à l'enseignement des mathématiques est venu compléter le questionnaire. Lors de la seconde édition de l'enquête en 2019, des questions relatives au lien avec les familles ont également été ajoutées.

## 5.2 Illustration et interprétation des écarts de vécus et de pratiques observés selon le réseau de scolarisation

Cette partie fournit une illustration des différences d'expériences et de pratiques observées selon le réseau de scolarisation et discute leur interprétation en s'appuyant sur les données collectées auprès des enseignants de CP en 2017/2018 lors de la première année de mise en œuvre de la réforme en REP+.

### 5.2.1 Comparabilité des répondants exerçant en REP+ et de ceux exerçant en REP (Proche REP+) et interprétation des écarts

Les données ont été collectées auprès de 3 290 enseignants de CP exerçant dans 1 131 écoles de REP+, 903 enseignants de CP exerçant dans 492 écoles de REP et 807 enseignants de CP exerçant dans 516 écoles publiques hors éducation prioritaire. Les taux de réponse sont d'environ 80 % et varient modérément selon le réseau de l'école (REP+, REP ou hors EP). Parmi l'échantillon de répondants, on compte 883 enseignants de l'échantillon « élèves » (dont 559 exerçant

en REP+, 174 en REP et 150 hors EP) <sup>11</sup>.

Les professeurs des écoles du niveau CP qui exercent en REP+ en 2017-2018 sont a priori bénéficiaires de la réduction de la taille des classes : ceux ayant répondu à l'enquête déclarent en moyenne une taille de classe de 12,8 élèves contre 20,6 élèves pour les enseignants exerçant en REP, ce qui confirme la mise en œuvre effective de la mesure.

Les données d'enquête ne permettent qu'une simple comparaison du vécu et des pratiques des enseignants des classes de CP dédoublées en REP+ avec celles des enseignants de CP en REP. En effet, la comparaison des réponses des enseignants de REP+ avec celles des enseignants du groupe de comparaison en REP (Proche REP+) ne permet pas d'isoler l'effet de la réduction de la taille des classes de CP en REP+ sur l'expérience d'enseignement et les pratiques pédagogiques mises en œuvre en classe. Des différences de composition des groupes d'enseignants en REP+ et en REP pourraient par exemple contribuer aux écarts observés entre les enseignants exerçant en REP+ et ceux exerçant en REP (les enseignants affectés aux classes dédoublées ont pu être « sélectionnés » en raison de caractéristiques particulières).

Le Tableau 7 montre notamment que les enseignants exerçant en REP+ ont moins d'ancienneté (-1,3 an sur le niveau CP) que leurs collègues de REP. Ils sont également plus nombreux à être diplômés d'un niveau Master, à être titulaires de leur poste ou à exercer à temps plein, mais les écarts observés sont de faible ampleur (2 à 5 points de pourcentage). Ces différences de composition restent donc modérées, mais l'on ne peut pas exclure que des différences inobservées compromettent la comparabilité des groupes et biaisent l'interprétation des écarts décrits dans l'analyse des pratiques. Dans les analyses des écarts de pratiques ou de ressentis entre REP+ et REP, on tiendra compte de ces différences en incluant les caractéristiques individuelles des enseignants pour lesquelles on observe des écarts de composition comme variables de contrôle dans les régressions. Notons, par ailleurs, que ces petits déséquilibres de composition des groupes REP+ et REP sont associés, pour l'échantillon d'enseignants exerçant hors EP, à des très petites variations des indicateurs synthétiques de pratiques (au plus de l'ordre, en valeur absolue, de 0,02 point des indicateurs non standardisés). Il faut également noter qu'il existe en REP+ une organisation de service spécifique permettant de consacrer davantage de temps pour le travail en équipe et le suivi des élèves. Les enseignants y exerçant bénéficient également de jours de formation continue supplémentaires. Les écarts observés entre les réponses des enseignants exerçant en REP+ avec celles de leurs collègues de REP doivent donc s'interpréter avec prudence.

---

11. Les résultats reportés dans cette partie s'appuient sur les données de l'échantillon élargi d'enseignants de CP, mais il convient de noter que les analyses présentées ont été dupliquées en restreignant l'échantillon aux seuls enseignants de CP de l'échantillon « élèves » ; elles donnent qualitativement les mêmes résultats que ceux observés pour l'échantillon élargi

TABLEAU 7 – Comparabilité des répondants selon le réseau d'éducation prioritaire

	Moyenne CP REP (Proche REP+)	Écart en CP REP+	N
% de femmes	85.9	-0.2 (1.4)	3 668
Âge moyen	40.6	-0.4 (0.3)	3 66
% pour lesquels enseigner est le 1 <sup>er</sup> choix de carrière	76.5	-1.3 (1.8)	3 658
% de diplômés d'un Master ou plus	36.1	5.8*** (1.9)	3 671
% de titulaires	93.8	2.0** (0.9)	3 681
% de temps plein	86.4	3.7*** (1.4)	3 573
Ancienneté moyenne, 1er degré	13.9	-0.9** (0.3)	3 663
Ancienneté moyenne, CP	6.5	-1.3*** (0.2)	3 651

Note : La significativité statistique est indiquée par les étoiles : \* pour une significativité à 10 %, \*\* à 5 % et \*\*\* à 1 %. La colonne N donne le nombre d'observations (celui-ci peut varier en fonction de la non-réponse à certaines questions de l'enquête).

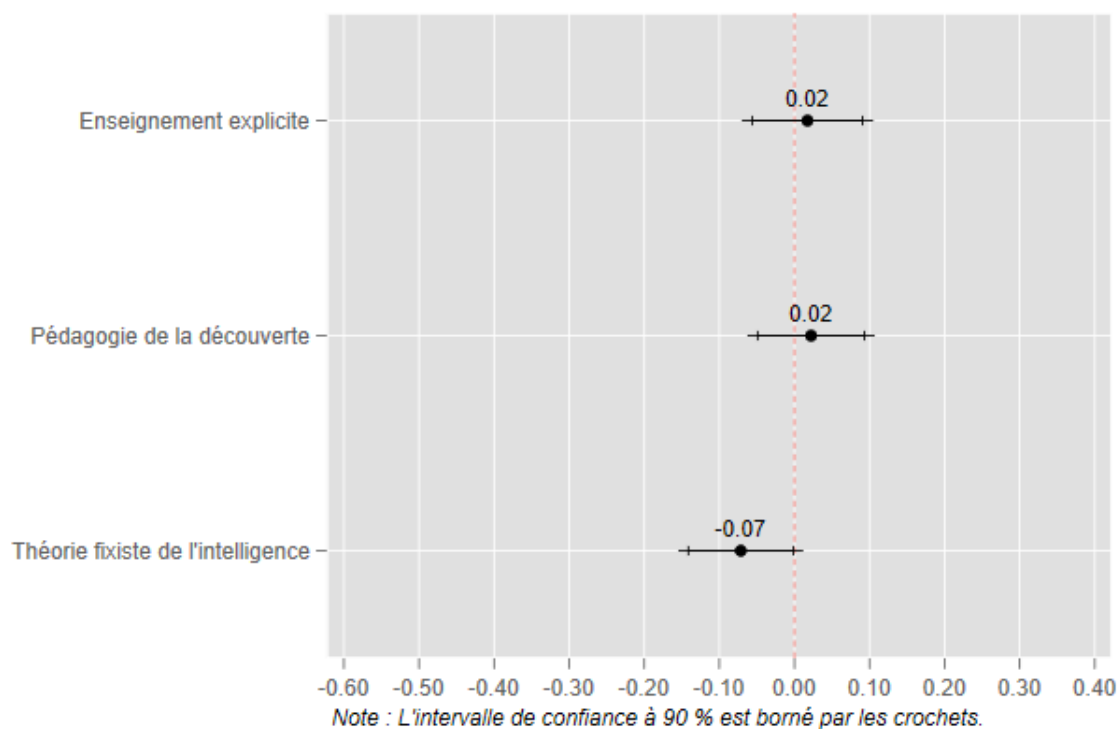
Champ : Enseignants de CP exerçant en REP+ et REP en 2017-2018 (échantillon de l'étude).

Source : MENJS-DEPP (enquête DEPP auprès des enseignants de CP, mai-juin 2018).

Comme indiqué précédemment les réponses aux questions portant sur les représentations des enseignants en matière d'enseignement et d'apprentissage peuvent également être utilisées pour tester la comparabilité des groupes d'enseignants de l'évaluation. Ainsi, un indicateur synthétique décrivant les opinions des enseignants en faveur de la pédagogie de la découverte peut être bâti à partir de la moyenne des items comme « *C'est lorsque les élèves résolvent eux-mêmes des problèmes qu'ils apprennent le mieux* », « *En tant qu'enseignant(e), mon rôle est d'aider les élèves à effectuer leurs propres recherches* » ou « *Il est préférable que l'enseignant laisse les élèves réfléchir eux-mêmes à des solutions pour résoudre des problèmes pratiques avant de leur montrer la marche à suivre* ». Plus la valeur de cet indicateur synthétique est élevée et plus l'enseignant adhère à des convictions pédagogiques orientées vers la pédagogie de la découverte. Cet indicateur synthétique est difficile à interpréter en valeur absolue ; il sert avant tout à comparer des groupes entre eux (ici des enseignants regroupés selon leur réseau d'éducation prioritaire, REP+ ou REP). Une pratique usuelle dans la littérature scientifique consiste à standardiser les indicateurs synthétiques par rapport à un groupe de référence, ce qui permet ensuite d'exprimer les écarts en pourcentages d'écart-type de l'indicateur synthétique du groupe de référence (dont la moyenne vaut 0 par construction). Dans l'analyse de l'enquête administrée aux enseignants de CP en 2017-2018, tous les indicateurs synthétiques construits pour les enseignants sont standardisés par rapport à ceux des enseignants de REP (ces derniers servant ainsi de point de référence). L'ampleur des écarts observés entre REP+ et REP pour chaque indicateur synthétique peut être comparée d'un domaine à l'autre, ce qui permet, par exemple, de décrire les indicateurs pour lesquels l'écart est le plus fort entre les enseignants de CP exerçant en REP+ et ceux exerçant en REP. S'agissant de l'indicateur synthétique décrivant l'adhésion des enseignants à des convictions pédagogiques tournées vers la pédagogie de la découverte, on ne constate pas d'écart statistiquement significatif entre le groupe des enseignants de REP+ et celui des enseignants de

REP (Figure 3).

FIGURE 3 – Écarts entre REP+ et REP relatifs aux convictions pédagogiques des enseignants de CP)



Lecture : S'agissant des convictions pédagogiques en faveur de l'enseignement explicite, l'écart estimé entre REP+ et REP correspond à 2 % d'un écart-type de l'indicateur synthétique en REP, mais il n'est pas statistiquement significatif : on le voit graphiquement, car l'intervalle de confiance de la différence entre REP+ et REP (ligne bleue horizontale encadrant l'estimation de l'écart représentée par le point bleu) inclut la valeur 0 de l'axe des abscisses. Lorsque l'intervalle des valeurs (pour l'estimation de l'écart) entre crochets n'inclut pas le 0, la significativité statistique est établie avec un niveau de confiance de 90 %. Lorsque c'est l'ensemble de ligne bleue (les valeurs comprises entre les crochets et celles à l'extérieur des crochets) qui n'inclut pas le 0, la significativité statistique est établie avec un niveau de confiance de 95 %. Tous les graphiques de la section se lisent de cette manière.

Champ : Enseignants de CP exerçant en REP+ et REP en 2017-2018 (échantillon de l'étude).

Source : Enquête DEPP auprès des enseignants de CP, mai-juin 2018.

Les enseignants ne se différencient pas non plus par leurs convictions pédagogiques à l'égard de l'enseignement explicite, c'est-à-dire un mode d'enseignement caractérisé notamment par une démarche directe, structurée et fortement guidée par l'enseignant. En REP+, comme en REP, environ un enseignant de CP sur deux est « d'accord » ou « tout à fait d'accord » avec l'affirmation selon laquelle « *En début d'apprentissage, il est important de guider fortement les élèves pour leur permettre de réussir les tâches proposées* ». Une proportion similaire considère que l'enseignement devrait être construit autour de problèmes qui ont des solutions claires et déterminées, et autour d'idées dont la plupart des élèves peuvent se saisir rapidement. De manière générale, les enseignants de CP, quel que soit leur réseau d'EP, rejettent la conception « fixiste » de l'intelligence qui reviendrait à considérer l'intelligence des élèves comme une capacité fixe et stable. Par exemple, 41 % des enseignants de CP en REP+ ne sont « pas du tout d'accord » avec l'affirmation selon laquelle « *Les résultats d'un élève à l'école sont quelque chose que l'on peut*

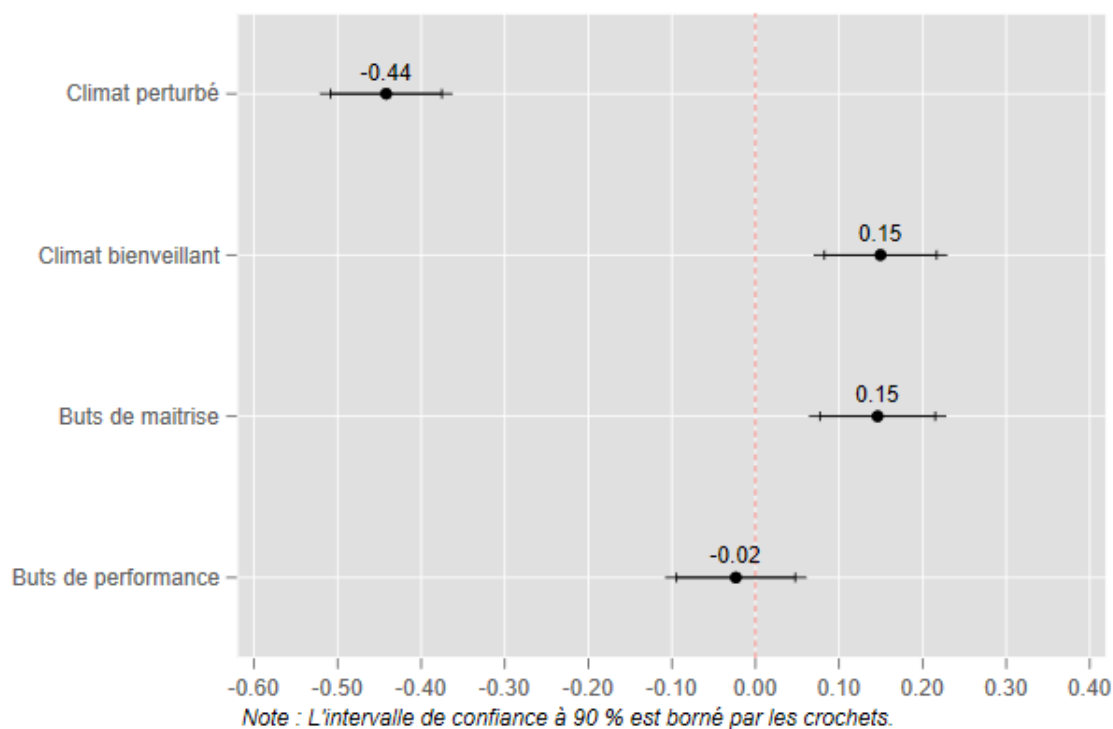


*difficilement faire évoluer* », contre 38 % en REP. S'agissant de l'indicateur synthétique décrivant l'adhésion à une conception fixiste de l'intelligence, l'écart entre le groupe des enseignants de REP+ et celui des enseignants de REP est à la limite de la significativité statistique.

### 5.2.2 Écarts de vécus et de pratiques des professeurs des écoles de CP en 2017-2018 entre REP+ et REP (Proche REP+)

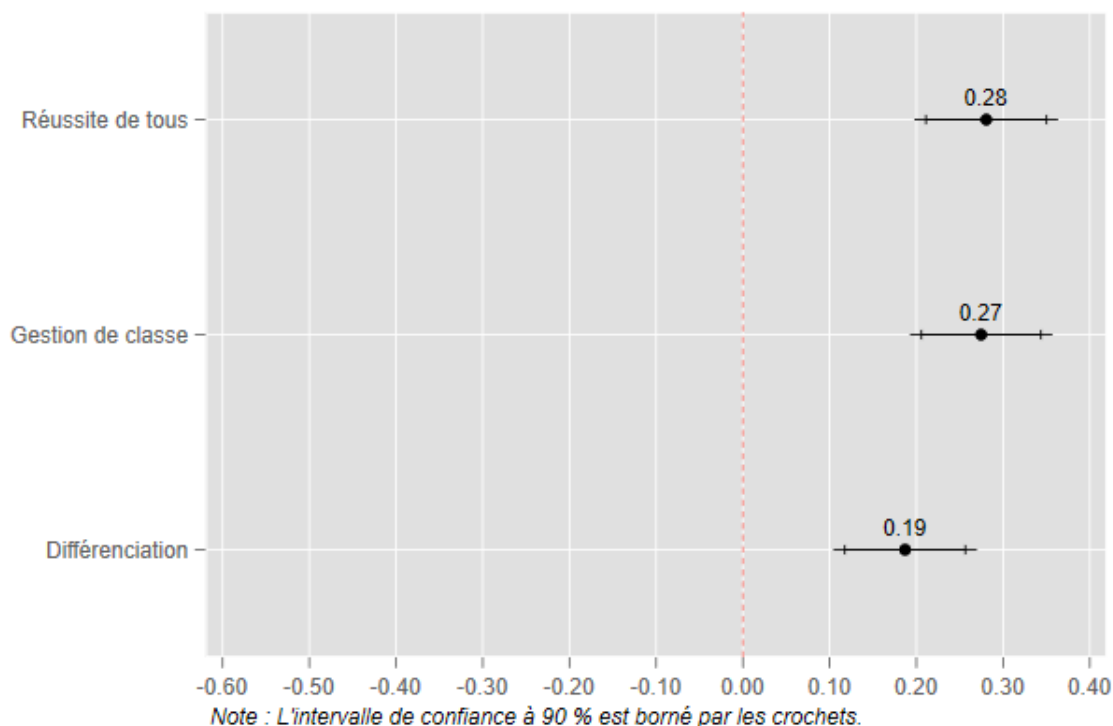
*Fonctionnement de la classe et sentiment d'efficacité personnelle.* En REP+, les enseignants de CP rapportent un climat de classe moins perturbé (-44 % d'un écart-type de l'indicateur synthétique en REP) (Figure 4) : par exemple, ils sont seulement 14 % (contre 21 % en REP) à déclarer être fréquemment interrompus par des bavardages et 21 % (contre 36 % en REP) à indiquer perdre « souvent » ou « toujours » du temps à cause d'élèves qui perturbent leurs séances. Cet écart doit être souligné car la gestion de classe est un enjeu majeur pour les enseignants français d'après TALIS. En outre, les enseignants de CP exerçant en REP+ ont davantage confiance en leurs aptitudes à gérer les comportements perturbateurs en classe (+27 % d'un écart-type de l'indicateur synthétique en REP) (Figure 5). Par exemple, 63 % d'entre eux rapportent parvenir dans une large mesure à amener les élèves à respecter les règles en classe, contre 57 % en REP. L'écart en points de pourcentage entre REP+ et REP est encore plus fort lorsque l'on considère la proportion d'enseignants rapportant une grande capacité à gérer les comportements perturbateurs en classe ou à calmer un élève bruyant ou perturbateur (respectivement, +11 et +12 points de pourcentage).

FIGURE 4 – Écarts entre REP+ et REP relatifs au fonctionnement des classes de CP)



Champ : Enseignants de CP exerçant en REP+ et REP en 2017-2018 (échantillon de l'étude).  
 Source : Enquête DEPP auprès des enseignants de CP, mai-juin 2018.

FIGURE 5 – Écarts entre REP+ et REP relatifs au sentiment d'efficacité personnelle des enseignants de CP)



Champ : Enseignants de CP exerçant en REP+ et REP en 2017-2018 (échantillon de l'étude).

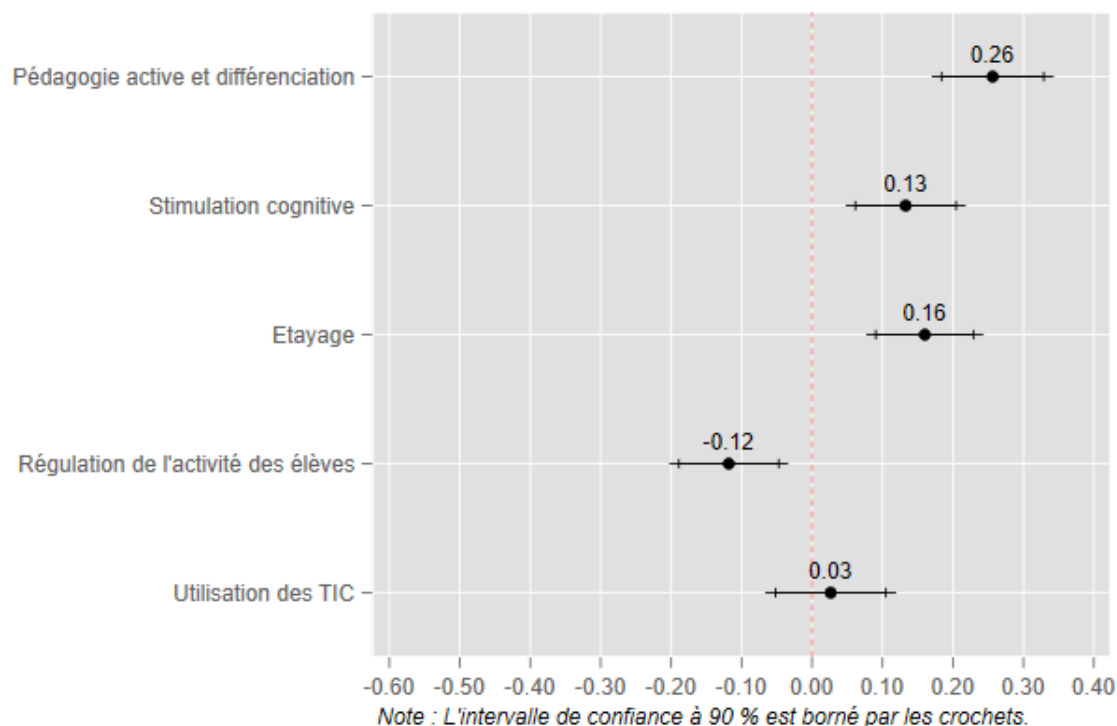
Source : Enquête DEPP auprès des enseignants de CP, mai-juin 2018.

Le climat motivationnel instauré par les enseignants en REP+ est davantage orienté vers des buts de maîtrise (+15 % d'un écart-type de l'indicateur synthétique en REP) (Figure 5.2). Ainsi, 51 % des enseignants en REP+ (contre 43 % en REP) considèrent que leurs pratiques correspondent « bien » ou « tout à fait » à la description suivante : « *Durant la classe, j'essaie d'offrir des activités différentes à faire, afin de permettre aux élèves de faire des choix* ». La quasi-totalité des professeurs des écoles en REP+ et en REP indiquent qu'ils essaient de donner aux élèves plusieurs exercices qui sont adaptés à leur niveau ou faire un effort particulier pour reconnaître les progrès individuels des élèves, même s'ils sont en difficulté. Les enseignants, dans l'ensemble, décrivent moins fréquemment un climat motivationnel tourné vers des buts de performance : par exemple, moins d'un enseignant sur trois, en REP+ comme en REP, considère que ses pratiques correspondent « bien » ou « tout à fait » à la description suivante : « *Je mets en valeur les élèves qui ont obtenu un bon résultat, comme un modèle à suivre* ». En outre, plus de 8 enseignants sur 10 (la proportion est similaire en REP+ et en REP) déclarent « toujours » complimenter leurs élèves quand ils font quelque chose de bien. Prises dans leur ensemble, les pratiques enseignantes en REP+ visent encore plus l'instauration d'un climat de classe bienveillant (+15 % d'un écart-type de l'indicateur synthétique en REP) : 70 % des professeurs des écoles de CP en REP+ (contre 65 % en REP) rapportent « toujours » indiquer à leurs élèves comment s'améliorer quand ils font une erreur ; 76 % (contre 69 % en REP) indiquent « toujours » avoir de bonnes relations avec les élèves. Ces écarts suggèrent que les enseignants de CP exerçant en REP+ cherchent davantage à motiver leurs élèves et à les soutenir dans leurs apprentissages.

Outre la dimension « gestion de classe », les valeurs des indicateurs synthétiques des enseignants de REP + relatifs aux autres domaines d'efficacité personnelle (« réussite de tous (les élèves) » et « différenciation ») sont supérieures à celles de leurs pairs en REP (respectivement, +28 % et +19 % d'un écart-type des indicateurs synthétiques en REP) (Figure 5). Ainsi, 51 % des enseignants de CP en REP+ rapportent parvenir dans une large mesure à appliquer différentes méthodes pédagogiques (*vs* 46 % en REP) ; ils sont 27 % à rapporter une grande capacité à expliquer les choses autrement en cas de difficultés de compréhension des élèves (contre 20 % en REP). Ils sont également 35 % à être « tout à fait d'accord » avec la proposition selon laquelle ils sont capables d'aider les élèves de leur classe à faire des progrès significatifs (*vs* 25 % en REP). De manière générale, les enseignants de CP exerçant en REP+ ont davantage confiance en leur efficacité à favoriser les apprentissages des élèves.

*Pratiques d'enseignement.* Outre le style motivationnel des enseignants, l'enquête les interrogeait à propos d'autres pratiques d'enseignement de façon générale ainsi que des méthodes liées à l'enseignement du français, et plus spécifiquement de la lecture. Dans les classes de CP en REP+, les professeurs des écoles se caractérisent davantage par des démarches d'enseignement qui recherchent la stimulation cognitive des élèves, avec l'intégration dans les pratiques de tâches stimulantes, nécessitant un raisonnement élaboré, favorisant la curiosité des élèves ou posant des questions amenant les élèves à réfléchir en profondeur (+13 % d'un écart-type de l'indicateur synthétique en REP) (Figure 6). Par exemple, ils sont 56 % à fréquemment proposer à leurs élèves des tâches qui nécessitent un raisonnement élaboré et 25 % à mettre fréquemment en place des tâches qui semblent difficiles à première vue (contre respectivement, 45 % et 18 % en REP). Leurs pratiques sont également plus actives et davantage orientées vers la différenciation. C'est sur cet indicateur de pratiques que la taille de l'écart observé entre REP+ et REP est la plus grande (+26 % d'un écart-type de l'indicateur synthétique en REP). Ainsi, les enseignants de REP+ sont 51 % à mettre fréquemment à disposition des élèves du matériel de manipulation, notamment pour les activités mathématiques (contre 37 % en REP). En REP+, les pratiques des enseignants mobilisent davantage des gestes professionnels se rapportant à un processus d'étayage pédagogique (+16 % d'un écart-type de l'indicateur synthétique en REP). Par exemple, ces derniers sont 75 % à montrer fréquemment aux élèves leurs progrès tout au long d'une activité (contre 66 % en REP, respectivement). À l'inverse, les enseignants de REP+ se caractérisent par un recours un peu moins fréquent à des gestes de régulation de l'activité des élèves (-12 % d'un écart-type de l'indicateur synthétique en REP) : par exemple, ils sont 77 % à donner « souvent », voire « à chaque séance ou presque » des indices si des élèves ne réussissent pas au bout d'un certain temps ou 79 % à simplifier « souvent », voire « à chaque séance ou presque » la consigne quand les élèves sont en difficulté (contre 79 % et 83 % en REP, respectivement). On ne note pas d'écart statistiquement significatif entre REP+ et REP s'agissant de l'indicateur synthétique « utilisation du numérique ».

FIGURE 6 – Écarts entre REP+ et REP relatifs aux pratiques générales d’enseignement des enseignants de CP)

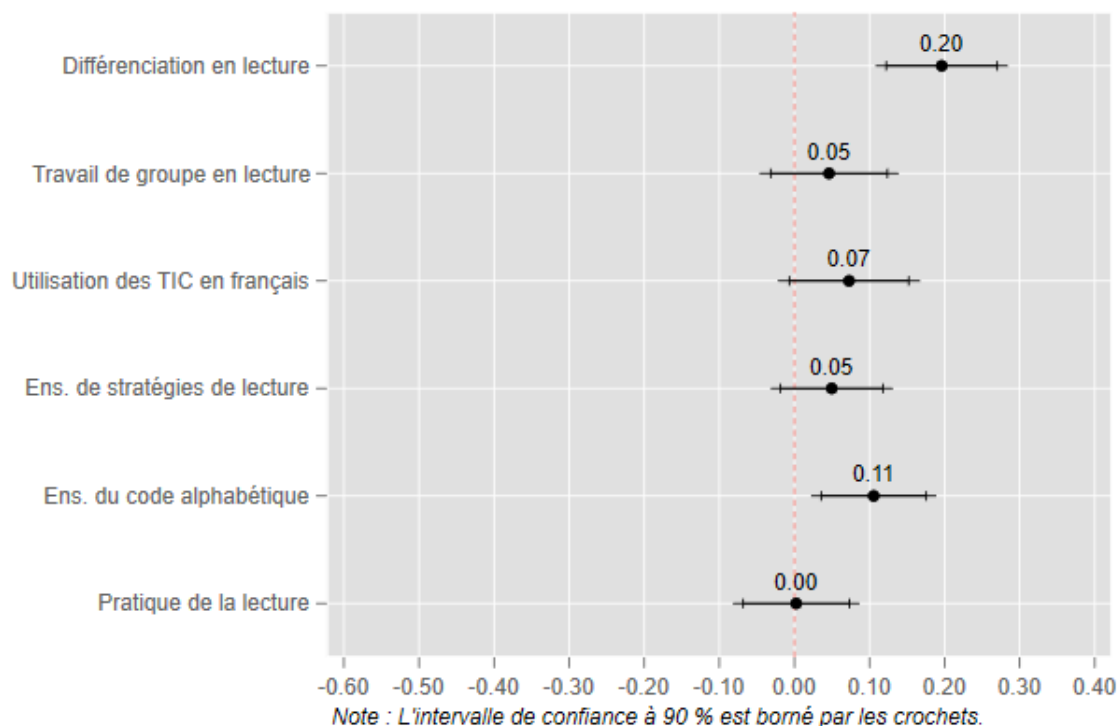


Champ : Enseignants de CP exerçant en REP+ et REP en 2017-2018 (échantillon de l'étude).

Source : Enquête DEPP auprès des enseignants de CP, mai-juin 2018.

L'enseignement de la lecture en REP+ s'appuie davantage sur la différenciation pédagogique (+20 % d'un écart-type de l'indicateur synthétique en REP) (Figure 7) : ainsi, 46 % des enseignants de CP de REP+ indiquent recourir fréquemment à l'enseignement individualisé pour la lecture (contre 38 % en REP). Ils sont également 36 % à rapporter que les élèves travaillent fréquemment seuls avec un plan ou un objectif à atteindre (contre seulement 30 % en REP). Ce résultat selon lequel la réduction de la taille de classe agit comme un levier pour la mise en œuvre de pratiques plus différenciées est sans surprise : un effectif de classe plus petit favorise en effet une meilleure prise en compte de la réalité individuelle de l'élève. L'enseignement de la lecture en REP+ s'appuie davantage sur l'enseignement du code alphabétique (+10 % d'un écart-type de l'indicateur synthétique en REP) : ainsi, 78 % des enseignants de CP de REP+ indiquent enseigner « chaque jour ou presque » le code alphabétique (contre 73 % en REP). Les professeurs des écoles de CP exerçant en REP+ ne se distinguent pas de ceux exerçant en REP s'agissant des autres indicateurs synthétiques relatifs à l'enseignement du français et de la lecture : travail de groupe en lecture, pratique de la lecture, enseignement de stratégies de lecture, utilisation du numérique en français.

FIGURE 7 – Écarts entre REP+ et REP relatifs aux pratiques d’enseignement du français des enseignants de CP)



Champ : Enseignants de CP exerçant en REP+ et REP en 2017-2018 (échantillon de l'étude).

Source : Enquête DEPP auprès des enseignants de CP, mai-juin 2018.

## 6 Observations de classe

Les observations de classe ont été menées aux printemps 2018, 2019 et 2021 auprès de 150 classes de CP de REP+, REP et hors EP, réparties dans 80 écoles appartenant à l'échantillon du volet « pratiques enseignantes ». Pour mettre en œuvre les observations de classe, cinquante conseillers pédagogiques, formés chaque année par le LaRAC à l'utilisation de trois outils d'observation, se sont rendus tous les ans, pour les écoles du dispositif CP12, deux jours dans chaque classe observée. Les conseillers pédagogiques travaillaient en binôme dans une même classe pour mener soit des observations simultanées sur trois élèves chacun, soit pour doubler la même observation des pratiques d'interactions de l'enseignant de la classe. Les trois outils d'observation utilisés sont :

1. Une grille d'observation de l'engagement comportemental des élèves dans les tâches scolaires ;
2. Une grille d'observation des interactions langagières ;
3. Un outil d'évaluation de la qualité des interactions, développé par Pianta, La Paro & Hamre (2008), le CLASS (Classroom Assessment Scoring System) de niveau K3.

Ces outils sont décrits plus en détail ci-dessous ; quelques résultats d'évaluation relatifs à la mesure de l'engagement des élèves dans les tâches scolaires et au style interpersonnel des enseignants, tirés de la mise en œuvre de ces outils, sont également fournis.

## 6.1 Outil 1 : grille d'observation de l'engagement comportemental des élèves dans les tâches scolaires

Deux observateurs par classe devaient mener une observation le matin et une l'après-midi. Chacun d'eux observait trois élèves alternativement toutes les dix secondes pendant dix minutes, soit vingt mesures d'engagement comportemental par élève. Le codage du comportement observable de l'élève à l'instant de l'observation pouvait prendre uniquement deux valeurs, qui étaient soit 0 lorsque l'élève était manifestement inattentif ou menait une activité étrangère à la tâche scolaire en cours, soit 1 lorsque l'attention de l'élève ou son activité semblait centrée sur la tâche en cours ou, a minima, n'était pas incompatible avec la tâche scolaire en cours. Au total, chaque période d'observation se solde par un score de 0 à 20 pour chaque élève. Cet outil d'observation a été développé par Altet, Bressoux, Bru et Lambert (1994) (Bressoux *et al.*, 1999).

Au cours des deux premiers temps d'observation (printemps 2018 et 2019), l'engagement de 1 721 élèves a été évalué au moins à une reprise, soit le matin, soit l'après-midi. On dénombre 1 681 élèves pour lesquels l'engagement a été mesuré à la fois le matin et l'après-midi au cours d'une même journée. Le Tableau 8 indique le nombre d'élèves observés le matin et l'après-midi en fonction de l'appartenance ou non à un réseau d'éducation prioritaire.

TABLEAU 8 – Nombre d'élèves observés à différents moments de la journée à l'aide de la grille d'observation de l'engagement

	2018		2019	
	Matin	Après-midi	Matin	Après-midi
REP+	505	500	474	474
REP (Proche REP+)	228	228	216	210
Hors EP	149	149	129	128

Champ : Échantillon de 150 classes de CP du secteur public.  
Source : MENJS-DEPP, LaRAC.

Selon Finn *et al.* (2003), la réduction de l'effectif devrait entraîner une plus forte implication des élèves dans les tâches d'apprentissage. Ainsi ils devraient moins souvent être occupés par des activités hors tâche scolaire, être plus souvent à l'initiative d'interactions avec l'enseignant. À l'inverse, on devrait observer moins d'interactions entre élèves car les élèves seraient moins dépendants de l'aide de leurs pairs, puisqu'ils peuvent plus bénéficier de l'aide directe de l'enseignant.

L'implication des élèves augmente-t-elle ? Pour répondre à cette question, il est possible d'analyser l'évolution de l'implication des élèves dans les activités scolaires à partir de la mesure de l'engagement comportemental. Lorsqu'on analyse l'évolution de l'engagement des élèves de l'éducation prioritaire, on constate qu'en REP+, où les effectifs avaient déjà été réduits en 2017-2018, l'engagement reste stable alors qu'en REP, où les effectifs ont été réduits en 2018-2019, l'engagement des élèves augmente significativement, passant d'une moyenne de 15,1 sur 20 mesures au printemps 2018 à une moyenne de 16,2 sur

20 au printemps 2019 et dépasse l'engagement moyen des élèves en REP+ la même année.

## 6.2 Outil 2 : grille d'observation des interactions langagières

Dans chaque classe, deux observateurs ont observé chacun trois élèves, pendant dix minutes par élève, par période de 30 secondes, soit vingt observations par élève. L'observateur observait l'élève pendant dix secondes puis, pendant les vingt secondes suivantes, codait ce qu'il avait observé. Deux catégories d'attributs de l'interaction étaient codées :

- Le type d'interaction, c'est-à-dire qui étaient les acteurs de l'interaction en distinguant six types possibles : interaction à l'initiative de l'enseignant vers l'élève observé ; interaction à l'initiative de l'élève vers l'enseignant ; interaction à l'initiative de l'enseignant vers plusieurs ou tous les élèves dont l'élève observé ; interaction de l'élève observé avec un ou plusieurs élèves ; interaction entre l'élève et un adulte autre que l'enseignant au sein de la classe (par exemple, lors d'un passage imprévu du directeur) ; pas d'interaction.
- L'objet de l'interaction, c'est-à-dire le contenu de l'interaction selon sept catégories. Ces catégories sont : « Contenu » lorsque l'interaction fait référence à des connaissances, à des activités scolaires ou au travail liés aux tâches scolaires ; « Organisation » lorsque l'interaction fait référence à l'organisation du travail, de la tâche ou du groupe ; « Comportement » lorsque l'interaction fait référence à la gestion des comportements ou à la discipline en classe ; « Social » lorsque le contenu de l'interaction porte sur la vie sociale ou la vie personnelle des interlocuteurs ; « Hors tâche » lorsque l'interaction ne faisait référence ni aux tâches et activités scolaires ni à la vie sociale de la classe ou des interlocuteurs ; la dernière catégorie « Indéterminé » était utilisée lorsque le contenu de l'interaction n'était pas identifiable.

Seules les interactions qui impliquaient l'élève observé étaient prises en compte pour le codage. Si, pendant la période des dix secondes d'observation, l'élève était impliqué dans plusieurs interactions, une seule était tout de même codée : celle qui apparaissait comme la plus signifiante du point de vue de l'élève et de ses apprentissages. Ainsi, les interactions individuelles primaient sur les interactions collectives et celles portant sur le contenu scolaire primaient sur celles portant sur les autres contenus. Au cours des deux premiers temps d'observation, les interactions de 1 682 élèves ont été observées à l'aide de cet outil. Le Tableau 9 indique le nombre d'élèves observés chaque année.

TABLEAU 9 – Nombre d'élèves observés à l'aide de la grille d'observation des interactions

	2018	2019
REP+	488	468
REP (Proche REP+)	225	223
Hors EP	150	128

Champ : Échantillon de 150 classes de CP du secteur public.  
Source : MENJS-DEPP, LaRAC.

Un autre indicateur de l'implication des élèves, fourni cette fois-ci par la grille d'observation des interactions langagières, peut être l'évolution des interactions à l'initiative des

élèves vers l'enseignant. On s'attendait à une convergence de leur fréquence pour les REP et REP+ en 2018-19 qui résulterait d'une stabilité de la fréquence de ce type d'interactions en REP+ puisque la taille des classes y était réduite dès 2017-2018, et d'une hausse de cette fréquence en REP puisque les classes n'y ont été réduites qu'en 2018-2019. On observe bien une tendance à la convergence mais, contrairement à cette hypothèse, elle est obtenue par une diminution, du nombre moyen d'interactions individuelles de chaque élève vers l'enseignant entre les deux années en REP+. Cette diminution est difficilement explicable. Signalons simplement à cet égard qu'il y a bien une hausse du nombre total d'interactions individuelles élève-enseignant et enseignant-élève en REP et que ce nombre converge vers celui documenté en REP+. C'est donc bien plutôt globalement à une hausse du nombre total d'interactions individuelles en REP qu'on assiste, passant de 1,58 au printemps 2018 à 2,01 au printemps 2019.

On s'attendait également à ce qu'ayant plus facilement accès à l'aide de l'enseignant, les élèves se tournent moins vers leurs pairs pour solliciter une assistance et donc qu'on observe moins d'interactions entre élèves. Comme supposé, on observe bien que la diminution des effectifs opérée en REP à partir de 2018-2019 s'accompagne d'une diminution des interactions entre élèves. Au printemps 2018, on en comptait en moyenne 2,47 quand on n'en observe plus que 1,77 au printemps 2019 soit un niveau identique à celui observé en REP+ l'année précédente (où ses interactions ont logiquement peu évolué entre les deux années scolaires).

### **6.3 Outil 3 : CLASS (Classroom Assessment Scoring System) de niveau K3**

Le CLASS est un outil d'évaluation de la qualité des interactions, développé par Pianta, La Paro & Hamre (2008). Il s'agit d'un outil d'observation et d'évaluation des interactions enseignant-élèves dans les trois domaines suivants : le soutien émotionnel, l'organisation et la gestion de classe, le soutien aux apprentissages (stratégies employées et mises en œuvre par l'enseignant dans la promotion des habiletés cognitives des élèves). Les domaines couverts par le CLASS sont décrits plus en détail dans le Tableau 10.

Chaque classe était observée quatre fois vingt minutes simultanément par deux observateurs qui devaient apprécier la qualité des interactions observées dans chaque dimension selon une échelle de 1 à 7 en s'appuyant sur le repérage ou l'absence d'indicateurs propres à chaque dimension. Le score final sur chaque domaine et dimension correspond à la moyenne de l'ensemble des cotations de ce domaine ou de cette dimension. Les scores peuvent donc évoluer entre 1, qualité faible, et 7, qualité élevée (sauf pour la dimension « Climat négatif de la classe » dont les scores sont inversés). Le Tableau 11 présente la structure des échantillons d'écoles et de classes observées en 2018 et en 2019 en fonction des réseaux de scolarisation. Comme pour les deux premiers outils d'observation, ces échantillons forment un sous-ensemble des échantillons des enquêtes adressées aux enseignants de CP. Parmi les enseignants observés en 2018 et 2019, 97 l'ont été les deux années. Parmi ces derniers, 52 exerçaient en REP+, 28 en REP et 17 hors éducation prioritaire.

Une synthèse des observations menées dans le monde à l'école primaire (entre l'équivalent de la moyenne section de maternelle et le CM2) utilisant cet outil d'observation a été produite par le Center for Advanced Study of Teaching and Learning. Elle indique



TABLEAU 10 – Description du CLASS

<b>Dimensions</b>	<b>Illustrations</b>
<i>Domaine 1 : Soutien émotionnel</i>	
Climat positif de la classe	Le respect, les relations positives, l'enthousiasme partagé
Climat négatif de la classe	Des manifestations d'affect négatif, de contrôle punitif, d'irrespect
Sensibilité de l'enseignant	La disponibilité de l'enseignant, son anticipation relative aux besoins, aux questions des élèves, à leurs difficultés
Prise en compte du point de vue de l'élève	La prise en compte de l'enseignant de l'intérêt, la motivation, et l'autonomie de ses élèves
<i>Domaine 2 : Organisation et gestion de la classe</i>	
Gestion des comportements des élèves	La clarté et le rappel des règles, la proactivité de l'enseignant face aux comportements des élèves
Productivité des élèves	L'organisation et la gestion fluide des activités, la préparation effective du cours
Modalités d'apprentissage	La variété et l'attractivité des modalités et du matériel, l'accompagnement de l'enseignant
<i>Domaine 3 : Soutien aux apprentissages</i>	
Développement de concepts	Faire comprendre plutôt que faire mémoriser, amener les élèves à raisonner, intégrer et développer des habiletés métacognitives
Qualité du feed-back de l'enseignant	Le feed-back sur la tâche, centré sur les processus, encourageant les élèves à aller plus loin dans leur réflexion
Interactions langagières	Le développement d'habiletés langagières, des conversations fréquentes

TABLEAU 11 – Effectifs observés à l'aide du CLASS

	2018			2019		
	REP		Hors	REP		Hors
	REP+	(Proches REP+)	EP	REP+	(Proches REP+)	EP
Écoles	37	22	21	37	22	18
Classes	82	38	25	80	36	22
dont enseignant déjà présent en 2018	s.o.	s.o.	s.o.	52	28	17
dont classes mononiveaux	79	32	17	80	36	12
dont classes multiniveaux	3	6	8	0	0	10
dont classes sans co-enseignement	78	38	25	65	31	22
dont classes à plusieurs enseignants (en co-enseignement)	5	0	0	15	5	0

Champ : Échantillon de 150 classes de CP du secteur public.  
Source : MENJS-DEPP, LaRAC.

qu'en moyenne, le domaine du soutien émotionnel est celui qui obtient les scores les plus favorables avec une moyenne proche de 5, et des scores qui sont considérés comme moyens à élevés. Le domaine de l'organisation et de la gestion de la classe est un peu moins bien évalué avec une majorité de scores qui se situent entre 4 et 5, donc dans une zone de qualité moyenne, enfin le domaine du soutien aux apprentissages est celui qui obtient les scores les plus faibles puisque la majorité des scores se situent entre 1 et 3 donc dans la zone de faible qualité. Les résultats observés dans notre échantillon de classes sont conformes à cette hiérarchie entre les trois domaines même si l'écart entre soutien émotionnel et organisation/gestion de la classe est moins marqué, les scores en matière d'organisation et de gestion de la classe et en matière de soutien aux apprentissages se situant dans la fourchette haute de ceux habituellement observés au niveau international (Tableau 12).

TABLEAU 12 – Scores moyens (et écarts-types) pour les trois domaines du CLASS dans l'échantillon de classes observées

	<b>Soutien émotionnel</b>	<b>Organisation et gestion de la classe</b>	<b>Soutien aux apprentissages</b>
2018	5,31 (0,961)	5,13 (1,008)	3,81 (1,202)
2019	5,28 (0,955)	5,22 (1,005)	3,93 (1,222)

Champ : Échantillon de 150 classes de CP du secteur public.

Source : MENJS-DEPP, LaRAC.

Selon Finn *et al.* (2003), la réduction d'effectif devrait s'accompagner d'une augmentation qualitative des interactions interpersonnelles de l'enseignant avec les élèves. On devrait également observer dans les classes à effectif réduit des interactions plus longues, notamment entre élèves. On devrait aussi constater une plus grande tolérance des enseignants à des comportements diversifiés de la part des élèves. On s'attend donc à observer en classe davantage d'interactions, liées au soutien émotionnel et au soutien aux apprentissages, dans les classes de CP à effectif réduit. En ce qui concerne l'organisation et la gestion de la classe, la situation pourrait être évaluée de façon plus contrastée puisque la plus grande tolérance des enseignants à la variabilité des comportements pourrait conduire à une évaluation moins favorable de la qualité de la gestion des comportements alors que, dans le même temps, la productivité des élèves pourrait s'améliorer, le temps dévolu à la gestion des comportements étant plus faible.

Pour vérifier ces évolutions attendues, il est possible de sélectionner dans l'échantillon, les enseignants de REP+ et REP présents en 2018 et en 2019. On pourra ainsi vérifier si la qualité des interactions au sein de la classe évolue plus lorsque l'effectif est réduit en cours de période d'observation, comme c'est le cas en REP pendant la période examinée. À noter qu'aussi bien en REP+ qu'en REP, si l'enseignant observé reste le même d'une année sur l'autre, ses élèves changent, ce qui peut aussi être source de variations, bien qu'on n'en connaisse guère l'ampleur. Pour étudier cette possible modification de la qualité des interactions chez les mêmes enseignants selon que l'effectif est réduit ou non, on commencera tout d'abord par examiner les corrélations entre années pour chaque dimension et domaine du CLASS pour les enseignants exerçant en REP+, où l'effectif des classes est resté stable, et pour ceux qui exercent en REP où l'effectif a été réduit la seconde année d'observation.

À travers l'analyse des corrélations entre les scores issus des observations de 2018 et 2019 en REP+ et REP, on constate une assez forte stabilité des scores sur la qualité des interactions en REP+ avec des corrélations significatives se situant toutes entre 0,51 et 0,67. En REP, les corrélations restent significatives et proches de ou supérieures à 0,50 pour le soutien émotionnel, la gestion de la classe, le climat positif, la gestion des comportements, la productivité des élèves et les modalités d'apprentissage. En revanche, les corrélations ne sont pas significatives pour les dimensions relatives au climat négatif, à la prise en compte de la perspective des élèves, au développement de concepts, à la qualité du feedback et qualité des interactions langagières, ainsi que pour le domaine du soutien à l'apprentissage qui agrège ces trois dernières dimensions. Il semble donc que, dans les classes de REP, qui ont vu une baisse d'effectif entre les deux mesures, l'évolution de la qualité des interactions ait plus tendance à fluctuer d'une année à l'autre pour un même enseignant, au moins sur une partie des dimensions, que dans les classes de REP+ où, rappelons-le, l'effectif est resté stable d'une année sur l'autre.

Dans les classes de REP, le niveau de corrélation élevé, avec des valeurs entre 0,51 et 0,58 dans les domaines du soutien émotionnel et de la gestion de la classe, tend à montrer que la qualité globale des interactions liées à la relation interpersonnelle et à la gestion pédagogique de classe reste assez constante chez un même enseignant malgré la modification du contexte entraînée par la diminution des effectifs. Il n'en va pas de même du domaine du soutien à l'apprentissage où l'on observe, sur les trois dimensions qui le constituent, des corrélations non significatives entre 2018 et 2019. Cette absence de corrélation semble être l'indicateur que l'effet de la réduction du nombre d'élèves sur les pratiques destinées à accompagner les élèves dans leurs apprentissages varie assez fortement d'un enseignant à l'autre. Ainsi, c'est dans le domaine du soutien à l'apprentissage que la réduction d'effectif viendrait le plus modifier les pratiques enseignantes. Lorsque l'effectif baisse, la qualité des interactions liées à ce domaine peut donc augmenter ou diminuer indépendamment de son niveau avant la baisse d'effectif.

En termes d'évolutions moyennes, malgré les perturbations identifiables par l'analyse des corrélations, on n'observe pas de modifications significatives des scores relatifs à la qualité des interactions en REP. En effet, une analyse de type ANOVA avec mesures répétées sur les écoles de REP+ et REP communes à 2018 et 2019 ne met en évidence aucune évolution significative (pas d'effet de l'année d'observation) ni d'évolution différente entre les REP+ dédoublées dès 2017-2018 et les REP dédoublées en 2018-2019 (pas d'effet d'interaction entre l'année d'observation et le réseau), et ce quel que soit le domaine ou la dimension du CLASS analysés. Il semble donc que la qualité des interactions en REP ne soit pas sensible, à court terme, à la diminution de l'effectif du nombre d'élèves, ce qui tendrait à montrer que les enseignants ne modifient pas significativement, de façon univoque, leurs pratiques lorsque l'effectif diminue. Le fait que les améliorations de la qualité des interactions attendues en REP ne soient pas observées et que l'analyse des corrélations indique d'assez fortes fluctuations de l'évolution de la qualité des interactions liées au domaine du soutien aux apprentissages d'un enseignant à l'autre pourrait être dû à la nécessité pour les enseignants de s'adapter à un nouveau contexte de classe. La première année de dédoublement pourrait alors être une année de transition pendant laquelle les enseignants devraient construire de nouvelles pratiques adaptées à une structure de classe différente du fait de la réduction des effectifs. Toutefois, si on observe des améliorations

significatives en REP+ lors de la seconde année d'exercice en classe à effectif réduit, il faut noter d'une part qu'elles sont de faible ampleur et, d'autre part, qu'elles sont cantonnées à certaines dimensions des domaines du soutien émotionnel (climat positif, sensibilité de l'enseignant) et de gestion de la classe (gestion des comportements). Cela semble indiquer que, même à plus long terme, la qualité des pratiques interactives des enseignants se modifie peu dans le domaine du soutien aux apprentissages.

L'autre source d'information sur la modification des pratiques enseignantes provient des interactions que les élèves reçoivent des enseignants. En effet, d'après Finn *et al.* (2003), celles-ci devraient baisser pour laisser plus de place aux interactions à l'initiative des élèves. Or, en REP, comme en REP+, on n'observe aucune variation significative du nombre total d'interactions collectives et individuelles qui sont adressées à chaque élève par l'enseignant entre 2018 et 2019. À nouveau, les observations ne mettent pas de modifications en évidence, d'un point de vue quantitatif, dans la façon dont les enseignants interagissent avec leurs élèves après la baisse d'effectif.

## Conclusion

La question de l'évaluation (et des défis qu'elle allait poser aux évaluateurs) de la politique de dédoublement des classes de CP et de CE1 en éducation prioritaire fut posée suffisamment tôt par rapport au déploiement de la politique pour permettre à la DEPP, avec l'appui de chercheurs spécialistes en sciences de l'éducation et en évaluation des politiques publiques, de concevoir et de déployer un protocole d'observation et d'évaluation statistique capable de produire des résultats fiables et pertinents pour éclairer l'action publique et nourrir le débat public sur les politiques éducatives. Pour ce faire, la DEPP a pu s'appuyer sur des données « administratives » tirées de ses systèmes d'information, mais n'a pu faire l'économie de collectes de données spécifiques afin de répondre à l'ensemble des questions d'évaluation. La richesse des données recueillies et des instruments de mesure utilisés offre par ailleurs des opportunités inédites pour conduire de nouvelles recherches avec d'autres objets d'étude, comme la mesure de l'effet enseignant en début de primaire ou une analyse de la comparabilité des résultats obtenus au travers d'enquêtes par questionnaire et d'observations de classe.

## Bibliographie

Altet, M., Bressoux, P., Bru, M., Lambert, C. (1994). Étude exploratoire des pratiques d'enseignement en classe de CE2, *Les Dossiers d'Éducation et Formations*, 44, Ministère de l'Éducation Nationale, DEP.

Andreu, S., Ben-Ali, L., Blouet, L., Bressoux, P., Charpentier, A., Cioldi, I., Lacroix, A., Lima, L., Murat, F., Odin-Steiner, D., Raffaëlli, C., Rocher, T., Vourc'h, R. (2021). « Évaluation de l'impact de la réduction de la taille des classes de CP et de CE1 en REP+ sur les résultats des élèves et les pratiques des enseignants », *Document de travail - série études*, n°2021.E04, septembre 2021, DEPP.

Angrist, J., Lavy, V., Leder-Luis, J., Shany, A. (2019). Maimonides' Rule Redux, *American Economic Review : Insights*, 1(3), 309-24.

Angrist, J., Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effects of Class Size on Scholastic Achievement, *The Quarterly Journal of Economics*, 114(2).

Bandura, A. (2010). Auto-efficacité. Le sentiment d'efficacité personnelle, Éditions de Boeck.

Blatchford, P. (2016). Is it true that class size does not matter. A critical review of research on class size effects. In P. Blatchford, K. W. Chan, M. Galton, K.-C. Lai, J. Chi-Kin Lee (Eds.), *Class size : Eastern and Western perspectives*, 92-104. London-New York : Routledge.

Blatchford, P., Bassett, P., Brown, P. (2005). Teachers' and pupils' behaviour in large and small classes : a systematic observation study of pupils aged 10/11 years, *Journal of Educational Psychology*, 97(3), 454-467.

Blatchford, P., Bassett, P., Brown, P. (2011). Examining the effect of class size on classroom engagement and teacher-pupil interaction : Differences in relation to pupil prior attainment and primary vs. secondary schools, *Learning and Instruction*, 21(6), 715-730.

Blatchford, P., Russell, A. (2020). Rethinking class size : The complex story of impact on teaching and learning. London : UCL Press. En ligne : <https://www.uclpress.co.uk/collections/education/products/166006>

Bohrnstedt, G.W., Stecher, B.M., Wiley, E.W. (2000). The California class size reduction evaluation : Lessons learned. In M.C. Wang & J.D. Finn (Eds.), *How small classes help teachers do their best*, 201-225. Philadelphia, PA : Laboratory for Student Success at Temple University Center for Research in Human Development and Education.

Bouguen, A., Grenet, J., Gurgand, M. (2017). La taille des classes influence-t-elle la réussite scolaire ?, *Les notes de l'IPP*, 28.

Bressoux, P. (1996). The effects of teachers' training on pupils' achievement : The case of elementary schools in France, *School Effectiveness and School Improvement*, 7(3), 252-279.

Bressoux, P., Bru, M., Altet, M., Lambert, C. (1999). Diversité des pratiques d'enseignement à l'école élémentaire, *Revue Française de Pédagogie*, 126, 97-110.

Bressoux, P., Kramarz, F., Prost, C. (2009). Teachers' training, class size and students' outcomes : learning from administrative forecasting mistakes, *Economic Journal*, 119, 540-561.

Bressoux, P., Lima, L. (2011). La place de l'évaluation dans les politiques éducatives : le cas de la taille des classes à l'école primaire en France. Raisons éducatives, gouverner l'éducation par les nombres ?, *Usages, débats et controverses*, 99-123.

Bryk, A. S. (2015). 2014 AERA distinguished lecture. Accelerating how we learn to

improve, *Educational Researcher*, 44(9), 467-477.

Buckingham, D. (2003). *Media Education : Literacy, Learning and Contemporary Culture*. Polity Press.

Cain, K., Oakill, J. (2007). *Children's comprehension problems in oral and written language. A cognitive perspective*. New York : Guilford press.

Charpentier, A., Embarek, R., Raffaëlli, C., Solnon, A. (2019). « Pratiques de classe, sentiment d'efficacité personnelle et besoins de formation : une photographie inédite du métier de professeur des écoles début 2018 », *Note d'Information*, n°19.22, MENJ-DEPP.

Chetty, R., Friedman, J., Hilger, N., Saez, E., Schanzenbach, D., Yagan, D. (2011). How Does your Kindergarten Classroom Affect your Earnings? Evidence from Project STAR, *The Quarterly Journal of Economics*, 76(4).

Dweck, C. S. (2006). *Mindset : The New Psychology of Success*. New York : Random House.

Evertson, C. M., Folger, J. K. (1989). Small class, large class : What do teachers do differently ? Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Folmer-Annevelink, E., Doolaard, S., Mascareño, M., Bosker, R. J. (2010). Class size effects on the number and types of student-teacher interactions in primary classrooms, *The Journal of Classroom Interaction*, 45(2), 30-38.

Finn, J. D., Pannozzo, G. M., Achilles, C. M. (2003). The "why's" of class size : Student behavior in small classes, *Review of Educational Research*, 73(3), 321-368.

Finn, J. D., Shanahan, M. E. (2016). Does class size (still) matter ? In P. Blatchford, K. W. Chan, M. Galton, K.-C. Lai, J. Chi-Kin Lee (Eds.), *Class size : Eastern and Western perspectives*, 121-145. London-New York : Routledge.

Fredriksson, P., Öckert, B., Oosterbeek, H. (2013). Long-Term Effects of Class Size, *The Quarterly Journal of Economics*, 128(1).

Galton, M., Pell, A. (2009). *Study on small class teaching in primary schools in Hong Kong*. Hong Kong : Education Bureau and Cambridge University.

Gough, P. B., Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1), 6-10.

Gurgand, M., Valdenaire M. (2012). « Le Fonds d'expérimentation pour la jeunesse et les politiques éducatives : premier retour d'expérience », *Éducation & formations*, n°81, pp 27-40.

Fougère D. (2012). « Les méthodes d'expérimentation en question », *Éducation & for-*

*mations*, n°81, pp 41-48.

Harfitt, G. J., Tsui, A. B. (2015). An examination of class size reduction on teaching and learning processes : A theoretical perspective. *British Educational Research Journal*, 41(5), 845-865.

Harfitt, G. J. (2012). An examination of teachers' perceptions and practice when teaching large and reduced-size classes : Do teachers really teach them in the same way? *Teaching and Teacher Education*, 28(1), 132-140.

Hattie, J. (2005). The paradox of reducing class size and improving learning outcomes. *International Journal of Educational Research*, 43 (6) 387-425.

Hattie, J. (2009). Visible learning : A synthesis of 800+ meta-analyses on achievement. London : Routledge.

Hattie, J. (2016). The right question in the debates about class size. Why is the (positive) effect so small? In P. Blatchford, K.W. Chan, M. Galton, K.-C. Lai J. Chi-Kin Lee (Eds.), *Class size : Eastern and Western perspectives*, 105-118. London-New York : Routledge.

Jepsen, C., Rivkin, S. (2009). Class Size Reduction and Student Achievement : The Potential Tradeoff between Teacher Quality and Class Size, *Journal of Human Resources*, 44(1).

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241-253.

Krueger, A. (1999). Experimental Estimates of Education Production Functions, *The Quarterly Journal of Economics*, 113(2).

Krueger, A. (2003). Economic Considerations and Class Size, *The Economic Journal*, 113(485).

Krueger, A., Whitmore, D. (2001). « The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results : Evidence from Project STAR », *The Economic Journal*, 111 (468).

Kyriakides, L., Christoforou, C., Charalambous, C. (2013). What matters for student learning outcomes : A meta-analysis of studies exploring factors of effective teaching, *Teaching and Teacher Education*, 36, 143-152.

Mingat, A. (1991). Expliquer la variété des acquisitions au cours préparatoire : les rôles de l'enfant, la famille et l'école, *Revue Française de Pédagogie*, 95, 47-63.

Monso, O. (2014). L'effet d'une réduction de la taille des classes sur la réussite scolaire en France : développements récents, *Éducation & Formations*, 85, 47-61.

Pianta, R. C., La Paro, K. M., Hamre, B. K. (2008). Classroom assessment scoring system : Manual K-3. Baltimore : Brookes.

Piketty, T. (2004). L'impact de la taille des classes et de la ségrégation sociale sur la réussite scolaire dans les écoles françaises. Une estimation à partir du panel primaire 1997. Paris School of Economics, mimeo.

Piketty, T., Valdenaire, M. (2006). L'impact de la taille des classes sur la réussite scolaire dans les écoles, collèges et lycées français. Estimations à partir du panel primaire 1997 et du panel secondaire 1995, *Les Dossiers-Enseignement scolaire*, 173, MEN-DEP.

Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Lepänen, P.H.T., Poikkeus, A.M., Tolvanen, A., Torppa, M., Lyytinen, H. (2007). Very early phonological and language skills : Estimating individual risk of reading disability, *Journal of Child Psychology and Psychiatry*, 48(9), 923-931.

Schmitt-Roland, S., Thaurel-Richard, M. (1997). Pratiques pédagogiques de l'enseignement du français en sixième et progrès des élèves, *Les dossiers d'Éducation & formations*, 87.

Thaurel-Richard, M., Verdon, R.. (1997). Pratiques pédagogiques de l'enseignement des mathématiques en sixième et progrès des élèves, *Les dossiers d'Éducation & formations*, 84.