



HAL
open science

À quoi servent les évaluations institutionnelles ?

Sylvie Plane, Catherine Brissaud

► **To cite this version:**

Sylvie Plane, Catherine Brissaud. À quoi servent les évaluations institutionnelles?. Repères : recherches en didactique du français langue maternelle, 2024, 69 (69), pp.7-35. 10.4000/1204s . halshs-04779917

HAL Id: halshs-04779917

<https://shs.hal.science/halshs-04779917v1>

Submitted on 11 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

À quoi servent les évaluations institutionnelles ?

What are institutional assessments used for?

Sylvie Plane
Sorbonne Université
Catherine Brissaud
Université Grenoble Alpes

Résumé : Après avoir évoqué les liens entre didactique du français et évaluation laissée à l'initiative des enseignants, cet article examine le rôle que l'on fait jouer aux évaluations institutionnelles, en particulier en francophonie. Les trois fonctions, certificative, sélective et informative de ces évaluations décidées par l'institution, fonctions parfois intriquées, sont rappelées. Elles sont ensuite détaillées en montrant, nombreux exemples à l'appui, comment on peut s'interroger sur leur évolution, parfois mouvementée, et sur ce qu'elles dévoilent du système éducatif et de son rapport au politique.

Abstract: After having mentioned the links between French language teaching and assessment left to the initiative of teachers, this article examines the role that institutional assessments play, particularly in the French-speaking world. These institution-initiated assessments' three certifying, selective and informative functions, which are sometimes intricate, are cited. They are then detailed, through a host of supporting examples, to illustrate how we can question their sometimes-eventful transformation and what they reveal about the education system and its relationship to politics.

Mots clés : évaluation, examen, politique éducative, enseignement

Keywords: evaluation, examination, education policy, teaching

Dans la vie sociale, *évaluer*, au sens étymologique d'estimer la *valeur* de quelque chose, fait partie des comportements ordinaires, spontanés, aux contours imprécis et dont les tenants et aboutissants peuvent échapper à leur auteur. Dans le domaine de l'éducation, l'acte d'évaluer est fonctionnel, finalisé et outillé. Il se définit, selon Charles Hadji (1989), comme un processus qui permet de vérifier la présence de quelque chose d'attendu, de situer un individu ou une production par rapport à une cible. Il constitue une source d'information qui éclaire la prise de décision aussi bien au niveau de la régulation des apprentissages par les enseignants dans leurs classes (Allal et Mottier Lopez, 2007) qu'à celui du pilotage des curriculums (Audigier *et al.*, 2006 ; Mottier Lopez, 2009 ; Mottier Lopez et Figari, 2012), et même au niveau des choix étatiques en matière de politique éducative (Mons, 2008). Son poids sur le destin des élèves explique que l'évaluation ait suscité de nombreux travaux, stimulés dans un premier temps, comme le rappelle Isabelle Delcambre (2007), par les recherches en docimologie qui se sont interrogées sur la notation, cette pratique

à laquelle l'évaluation est souvent assimilée et qui est si courante qu'on en oublie l'étrangeté alors qu'elle consiste à convertir en chiffres des données qualitatives.

Il n'est pas étonnant que la réflexion sur l'évaluation ait accompagné l'émergence des didactiques dans la mesure où les opérations en jeu dans l'évaluation – en particulier le fait de comparer une performance à des attendus, et en conséquence l'obligation de déterminer ces attendus – interrogent les disciplines et leur enseignement. La didactique du français s'est investie très vite, très activement et sans discontinuer, dans des recherches portant sur l'évaluation, sans doute parce que les objets d'enseignement du français sont extrêmement complexes, qu'il s'agisse de la lecture et de la production d'écrits (Petitjean, 1984 ; Groupe EVA, 1991) ou encore de l'oral (Garcia-Debanc, 1999 ; Nonnon, 2005) ou de la maîtrise de la langue (Nonnon, 2008 ; Kingsbury et Tremblay, 2008). L'intérêt pour cette question n'a jamais faibli, de très nombreux travaux ont été consacrés à l'étude des annotations des copies, à l'examen des ressources à disposition des enseignants en matière d'évaluation, aux problèmes que pose l'évaluation de l'oral, à la conception de dispositifs d'évaluation (Van Beveren *et al.*, 2013 ; Daunay et Dufays, 2014 ; Doquet et Pilorgé, 2020 ; Lafontaine *et al.*, 2016 ; Mauroux et Garcia-Debanc, 2013 ; Duvin-Parmentier *et al.*, 2021 ; Gagnon *et al.*, 2014 ; Gagnon *et al.*, 2019 ; Menouar et Chabanne, 2022 ; Nonnon, 2011 ; Falardeau *et al.*, 2016).

Mais ces évaluations ne sont pas des actes isolés, hors sol. À côté des processus mis en place par les enseignants à des fins d'information et de régulation qui leur sont propres, il existe des évaluations institutionnelles, notamment sous forme d'examens ou d'enquêtes, dont les enseignants n'ont pas l'initiative et qui non seulement leur fournissent des modèles mais sont sources d'obligations. Les enseignants sont en effet formellement tenus d'y préparer leurs élèves ou étudiants lorsque l'épreuve d'évaluation clôt leur période d'enseignement, ou, si leur classe n'est pas directement concernée, d'en tenir compte en raison de la publicité donnée aux résultats de ces évaluations.

À priori, trois fonctions sont réservées à ces évaluations institutionnelles :

- une fonction certificative, lorsqu'elles ont pour objet de mesurer les compétences d'apprenants considérés dans leur singularité et de permettre l'attribution d'un titre ou de contribuer à son attribution, comme l'examen suisse de maturité ou l'épreuve anticipée de français (EAF) du baccalauréat en France ;
- une fonction sélective, lorsqu'elles ont pour objet de mesurer les compétences d'individus mis en compétition, de comparer leurs performances et de les classer ; les concours de recrutement des enseignants en France en sont les exemples types ;
- une fonction informative, lorsqu'elles ne visent qu'à rendre compte d'un état, ce qui est la vocation affichée des tests standardisés à différents niveaux de scolarité, des évaluations-bilans

CEDRE¹, du dispositif PISA, des enquêtes d'utilité publique comme l'enquête Information et vie quotidienne de l'INSEE sur la littératie.

Ces fonctions ne sont ni incompatibles ni inamovibles. Ainsi, si leur méthodologie le permet, les tests nationaux que passent les élèves à certains moments de leur scolarité peuvent à la fois donner lieu à un traitement statistique par le ministère de l'Éducation et servir à prendre des décisions concernant les élèves individuellement.

Mais ces évaluations ne servent pas uniquement à produire des connaissances sur les savoirs et les compétences des apprenants : le pouvoir contraignant dont elles disposent et la place qu'elles occupent dans l'opinion publique permettent de les utiliser pour piloter l'action des enseignants et font d'elles des instruments majeurs des politiques éducatives.

Les enjeux considérables auxquels sont associées les évaluations institutionnelles expliquent que ce numéro de *Repères* ait pour objet de s'interroger sur les rôles multiples qu'elles jouent, sur l'évolution de ces rôles et sur la façon dont elles infléchissent l'enseignement du français. En quelque sorte, il s'agit, à partir de l'analyse d'évaluations portant sur des points particulièrement délicats de l'enseignement du français, de se demander non seulement à quoi servent les évaluations institutionnelles, comme nous le posions en titre, mais aussi de s'inquiéter de ce à quoi on les fait parfois servir.

1. Des évaluations à fonction certificative pour réguler l'enseignement

L'établissement d'évaluations institutionnelles a été, en France comme dans beaucoup d'autres pays, une lente conquête de la démocratie. L'institutionnalisation de l'école commence en France avec la loi de 1806 qui crée l'Université impériale, mais c'est avec la naissance de l'école républicaine dans les années 1880 que l'évaluation devient une préoccupation institutionnelle. Même si le mot « évaluation » ne constitue pas une entrée dans le dictionnaire de Buisson – il n'est employé à cette époque que dans le vocabulaire de l'économie – l'évaluation fait partie des préoccupations de ce grand pédagogue qui a mené bataille autour de la réglementation des examens pour dégager l'enseignement du français de sa soumission exclusive à l'apprentissage de l'orthographe et lui donner plus de corps (Chervel, 1995, p. 35). Pour tenter d'empêcher les dictées fourmillant de subtilités orthographiques d'écraser les autres épreuves, Ferdinand Buisson écarta les recteurs du choix des sujets du brevet et plaça cet examen, dont il fit la porte d'entrée à l'École normale, sous sa propre responsabilité. Il obtint que l'examen du certificat d'études, institué

¹ Le CEDRE (Cycle des évaluations disciplinaires réalisé sur échantillon) mis en place en France depuis 2003, mesure chaque année les performances dans un champ disciplinaire donné des élèves en fin d'école primaire (classe de CM2) et en fin de collège (classe de 3^e). La dernière évaluation dans le champ du français a eu lieu en 2021.

comme clôture de l'enseignement primaire par la loi de 1882, comprenne une épreuve de rédaction, ce qui introduisit officiellement à l'école primaire l'enseignement de la rédaction que ne prévoyaient pas jusqu'alors les plans d'études. Brevet et certificat d'études ont donc servi à la modernisation et à l'homogénéisation de l'enseignement en faisant bouger les équilibres entre les composantes disciplinaires et en désignant de nouveaux objets d'enseignement.

Ces mesures qui ont accompagné l'instauration de l'école républicaine montrent à l'évidence que les évaluations institutionnelles à fonction certificative, situées en fin d'un cycle d'études, ont une portée qui s'étend bien en amont de l'année où se situent les épreuves, et elles constituent en cela des outils de pilotage de l'enseignement et plus largement du système éducatif.

1.1. Les évaluations certificatives : des outils pour maîtriser les évolutions sociétales ou pour les accompagner ?

Les évaluations qui sanctionnent la fin d'une période de scolarité ont une dimension politique dans la mesure où elles sont impliquées dans la régulation des flux d'élèves et peuvent servir à maintenir des clivages sociaux par des exigences inaccessibles à une partie de la population, ou, à l'inverse, soutenir des changements en favorisant l'émergence de diplômés parmi de nouvelles catégories de populations. Elles constituent donc en quelque sorte des charnières qui peuvent ouvrir ou non vers une nouvelle période.

Le baccalauréat en France est un bon exemple de la manière dont les examens et le contexte social et politique ont partie liée, et son évolution en témoigne. Le baccalauréat a en effet longtemps servi à conforter un certain entre-soi, à une époque où les carrières devaient plus à la fortune et aux relations qu'aux études. Sa création, qui a suivi de quelques années celle du lycée, s'inscrit dans le cadre de la mise en place d'un système scolaire qui, pour maintenir la stabilité sociale, organise deux parcours parallèles, celui de l'enseignement primaire et celui de l'enseignement secondaire, destinés à « reconstruire la société française sur une base de classe » (Chervel, 1993, p. 35). Le latin a eu sa part dans cette politique. En effet, le baccalauréat, dont la première session en 1809 a permis à 31 candidats d'être diplômés (Meuriot, 1919), et qui a consisté jusqu'en 1840 en une épreuve orale, a exigé jusqu'en 1880 la connaissance du latin. Or le latin n'était enseigné que dans les lycées – payants et réservés aux garçons – ou dans les institutions religieuses pour garçons, si bien que ni les élèves des collèges communaux, ni ceux de l'enseignement primaire supérieur, ni les filles ne pouvaient se présenter à cet examen dont la fonction principale était, selon Bruno Belhoste (1990), de rassurer les élèves formés à la culture classique sur la « certitude intime de [leur] distinction sociale ». L'introduction en 1881 d'une composition à rédiger en français, exercice considéré comme « non noble » (Chervel, 2006, p. 598), constitue une timide promotion de la discipline « français », mais il faudra attendre 1902 pour qu'il soit possible de passer un baccalauréat « moderne », c'est-à-dire sans latin. Toutefois, les études de droit et de médecine refuseront de s'ouvrir

aux titulaires de ce diplôme. Les filles attendront 1919 pour avoir l'autorisation de passer un baccalauréat qui leur est réservé² et 1924 pour que les programmes soient communs aux deux sexes. La croissance du nombre de diplômés sera lente puisqu'en 1951, avec 33 000 lauréats (Comité national d'évaluation, 1987), le baccalauréat ne concerne que 5 % d'une classe d'âge. Il n'atteint 20 % qu'en 1980 avec 160 000 bacheliers. Ce n'est qu'à partir de cette date qu'il connaîtra une véritable expansion, grâce à une politique volontariste et à la création du bac professionnel. Cette politique a porté ses fruits puisqu'en 2022 le baccalauréat a été décerné à 664 300 candidats, soit 80 % d'une classe d'âge, avec un taux global de réussite de 91,1 % et de 96,1 % dans la voie générale. Mais la démocratisation puis la massification ont eu pour corolaire la dévalorisation du diplôme, malgré ses exigences (Maillard, 2021).

Même s'il n'a conservé que sur le plan juridique, et non dans les pratiques, sa reconnaissance comme premier grade de l'université, le baccalauréat est un diplôme qui non seulement valide la scolarité secondaire mais constitue également une clé d'entrée vers les études supérieures. La situation de l'examen de maturité qui, en Suisse, sanctionne la fin des études secondaires post-obligatoires est relativement différente, parce que le contexte est plus complexe – nous y reviendrons – et parce que cet examen ouvre plus ou moins de portes selon le cadre dans lequel il est passé. En effet, si les certificats de maturité gymnasiale préparés dans les gymnases cantonaux ou reconnus par les cantons donnent tous accès aux hautes écoles universitaires suisses, malgré des spécificités cantonales, en revanche, seule la maturité fédérale (appelée aujourd'hui *examen suisse de maturité*), qui se prépare à partir de l'âge de 18 ans dans des établissements privés généralement coûteux, donne également accès aux universités étrangères.

Les évaluations certificatives qui interviennent en cours de scolarité sont moins prestigieuses que celles qui, comme le baccalauréat, couronnent la totalité du cursus secondaire. Pourtant, elles peuvent avoir un poids considérable si leur réussite conditionne l'accès au cycle suivant, ce qui est le cas du certificat d'études de base (CEB) que les écoliers de la Communauté francophone de Belgique doivent passer pour accéder à l'enseignement secondaire. Une fois ce seuil franchi, c'est l'établissement secondaire où est scolarisé l'élève qui octroiera le *quitus* ouvrant l'accès à l'université en lui délivrant le certificat d'enseignement secondaire supérieur (CESS).

En revanche, en France, l'examen appelé depuis 1988 diplôme national du brevet (DNB) que les élèves passent à la fin du collège ne conditionne pas l'entrée au lycée, du moins jusqu'à présent³. Il n'a pas non

² Le cas de Julie-Victoire Daublé qui obtint le baccalauréat en 1861 est resté longtemps isolé. À noter qu'elle parvint également à obtenir une licence, alors que les cours de la Sorbonne n'étaient pas ouverts aux femmes.

³ Le dossier de presse de la conférence ministérielle du 5 décembre 2023 indique que le DNB conditionnera l'entrée au lycée, mais le site du ministère de l'Éducation nationale, consulté le 3 avril 2024, mentionne toujours :

plus de valeur sur le marché de l'emploi et est donc associé à peu d'enjeux, mais il atteste cependant des connaissances acquises en fin de scolarité obligatoire et est, depuis 2005, explicitement relié au socle de compétences, de connaissances et de culture. Actuellement, la partie examen, qui représente la moitié de la note finale, n'est réellement déterminante que pour les élèves qui ont obtenu un faible nombre de points au contrôle continu (évaluation du socle), mais qui sont en même temps ceux qui risquent d'avoir le DNB pour seul diplôme. Sans surprise, le résultat à l'écrit, et plus largement l'obtention du DNB sont fortement corrélés à l'origine sociale des élèves : à la session 2023, le taux de réussite s'élevait à 99 % dans les établissements privés sous contrat, qui accueillent des élèves favorisés (indice médian de position sociale⁴ : 116), et n'atteignait que 81 % dans les établissements situés en REP+⁵ (IPS médian : 74). Or, le ministre français de l'Éducation a annoncé le 5 décembre 2023 que le passage au lycée serait conditionné par l'obtention du DNB, et que les élèves qui ne l'auraient pas obtenu seraient scolarisés dans des classes spécifiques où ils prépareraient à nouveau ce diplôme. On peut craindre que les concepteurs du projet n'aient pas anticipé l'impact d'une telle mesure sur les lycées professionnels dont les effectifs risquent de baisser car ils sont majoritairement composés d'élèves ayant eu une scolarité difficile. La note de service et le décret n° 2024-229 du 16 mars 2024 qui opérationnalisent la décision du ministre, appelé depuis lors à d'autres fonctions, indiquent que les lycées où se situeraient ces classes spécifiques appelées « prépa-2^{de} » restent à identifier et donnent à penser que l'accueil des élèves concernés se ferait sur la base du volontariat, d'où probablement une montée en puissance de l'entrée directe en apprentissage.

1.2. Les évaluations certificatives entre harmonisation imposée et harmonisation à construire

De même que les programmes scolaires ont des pouvoirs prescriptifs inégaux en force et en précision selon les pays, les examens terminaux ont une fonction d'harmonisation de l'enseignement plus ou moins puissante selon les spécificités du système éducatif. Ainsi en France, même si l'obtention du baccalauréat ne constitue plus un exploit en raison du nombre élevé de lauréats chaque année, la préparation de ce diplôme continue de régir tous les enseignements dispensés aux 2 240 000 adolescents scolarisés en lycée, soit près de 85 % d'une classe d'âge (Dauphin *et al.*, 2022). Mais cette harmonisation s'exerce en trompe-l'œil, car comme le notent Vincent Dupriez et Marie Verhoeven (2013), le long processus d'unification des cursus auquel se sont attelés plusieurs pays européens, concrétisé en Belgique par la loi de 1969 et en

« L'attribution du brevet ne conditionne pas l'accès à une classe supérieure en fin de troisième » (voir <<https://www.education.gouv.fr/le-diplome-national-du-brevet-10613>>).

⁴ L'indice de position sociale (IPS), établi à partir des catégories socioprofessionnelles des familles, renseigne sur les caractéristiques de la population d'un établissement scolaire. Plus il est élevé, plus le milieu est favorisé.

⁵ Réseau d'éducation prioritaire renforcé.

France par les lois de 1975 (loi Haby) et 1985 (création des baccalauréats professionnels) s'est doublé de la mise en place d'un système complexe de ramifications et d'options qui rétablit de la différenciation : l'existence des voies générale, technologique et professionnelle, de spécialités et d'options fait que le même vocable « baccalauréat » désigne des parcours différents.

Les examens qui sanctionnent la fin d'un cycle ont néanmoins un rôle unificateur, et celui-ci peut même être capital dans les pays dont le système éducatif est moins férocement centralisé que celui de la France parce qu'ils définissent des attentes qui devront être partagées par tous les enseignants d'un même niveau, quel que soit le cadre dans lequel ils exercent. C'est ce que fait, par exemple, dans la Fédération Wallonie-Bruxelles le certificat d'études de base, évoqué plus haut, qui donne accès au secondaire et qui se substitue depuis 2009 à l'examen que les élèves passaient dans le cadre cantonal ou diocésain, selon qu'ils étaient scolarisés dans l'enseignement public ou dans l'enseignement catholique. Cet examen, que le livret scolaire peut compenser, vise à vérifier que tous les élèves ont les mêmes acquis dans les domaines disciplinaires – principalement le français et les mathématiques – et en termes de compétences, dans un contexte où la majorité des 500 000 écoliers se répartit entre deux réseaux d'enseignement distincts, le réseau officiel et le réseau catholique (195 000 écoliers).

Même s'ils ne se concrétisent pas par un dispositif qui ferait consensus, les processus visant à la mise en place d'évaluations certificatives sont féconds par les débats et dialogues qu'ils suscitent.

Ainsi, même si le panorama des examens en Suisse reste d'une rare complexité, avec quantité de combinaisons possibles en termes de calendrier et de contenu (Françoise Landry [2006] parle pour la Suisse romande et le Tessin de « huit systèmes scolaires cantonaux »), il a nourri des réflexions que se sont parfois épargnées les pays où les circuits de décision sont très descendants. La complexité des dispositifs suisses d'évaluation tient au peu d'appétence de ce pays pour la centralisation, même si la culture politique n'est pas exactement la même dans les cantons alémaniques, très soucieux d'autonomie, et dans les cantons francophones qui ont tendance à se référer davantage au niveau national (Mueller et Dadanelli, 2014). Avec quatre langues officielles, des cantons qui disposent d'une administration qui leur est propre, une absence de synchronisation fédérale des cycles scolaires, parvenir à un protocole unique était difficilement envisageable, mais le travail d'harmonisation a été riche. Sur le plan didactique, à l'initiative de l'Université de Genève, les examens de maturité de français langue première ont été tout particulièrement scrutés, afin de saisir ce qui relève de l'autonomie locale et ce qui peut être cadré et harmonisé (Monnier-Silva *et al.*, 2020). Sur le plan plus général de l'organisation des cursus, la Conférence suisse des directeurs cantonaux de l'instruction publique (CDIP) a contribué à l'harmonisation en définissant les compétences à faire acquérir aux élèves du secondaire dans l'optique d'une poursuite d'étude universitaire (Leuenberger et Sartori, 2015). Cette réflexion était nécessaire, car contrairement

aux pratiques d'externalisation des examens fréquentes dans beaucoup de pays, les certificats cantonaux de maturité gymnasiale sont délivrés dans l'établissement où sont scolarisés les élèves.

La régulation des évaluations institutionnelles qui a été opérée au Québec s'est déroulée à peu près à la même période que celles qui se déroulaient en Europe et a participé de la même aspiration à la démocratisation de l'enseignement. Mais elle a été également associée à des enjeux spécifiques, celui de la résistance à la domination culturelle anglophone, et celui de la laïcisation de l'enseignement dans un contexte où les structures religieuses avaient de grandes responsabilités en matière d'enseignement (Plumelle, 2002). L'entreprise était d'autant plus ardue que le réseau catholique avait été l'artisan de la préservation du français dans une province isolée au milieu d'un univers anglophone. L'harmonisation des dispositifs d'évaluation n'a pu se faire que grâce à la réduction du nombre de commissions scolaires, qui étaient 72 au départ et avaient une grande autonomie, puis à leur remplacement en 2020 par des centres de service scolaire, à l'exception des commissions anglophone, cri et inuktitut, qui sont toutes trois maintenues pour des raisons linguistiques. Cette politique est à rapprocher de celle qui a permis au Royaume-Uni de contrecarrer l'influence des autorités religieuses qui avaient la main sur la supervision des établissements scolaires, jusqu'à ce que la création de l'Oftsed (Office for Standards in Education) en 1992 instaure un système national d'évaluation qui les a privées de leur possibilité de promouvoir des chaînes d'écoles et de collèges privés en se portant garantes de la qualité et de l'exigence des évaluations qu'elles pratiquent.

1.3. Suivre l'évolution des évaluations certificatives pour suivre l'évolution de l'enseignement

L'analyse en diachronie des épreuves donne à voir comment elles participent du figement ou au contraire de la transformation d'une discipline. L'évolution du DNB (diplôme national du brevet), que nous avons évoqué plus haut, est intéressante parce que cet examen clôt la période d'enseignement d'un certain nombre d'objets qui seront éventuellement évoqués les années suivantes mais ne seront plus étudiés spécifiquement. En effet, la classe de troisième, qui se termine par cet examen, est la dernière année de l'enseignement de la grammaire et de l'orthographe (elles feront uniquement l'objet de révisions et d'approfondissement les années suivantes). Elle est aussi la dernière année d'un cycle pour lequel les programmes de 2015, révisés plusieurs fois depuis leur publication, prévoient un enseignement de l'expression et de la compréhension orales et mentionnent explicitement parmi les compétences à développer celles relatives à la lecture de textes non littéraires tels les articles de presse.

Le DNB n'est pas exactement un examen de fin de scolarité obligatoire – laquelle est fixée à 16 ans en France – dans la mesure où les élèves passent cet examen lors de leur dernière année de collège, en principe à 15 ans, qui est aussi l'âge auquel se passe l'évaluation du PISA. Mais à l'heure où les résultats

du PISA pour la France baissent et où la fonction du DNB est en passe de changer (comme précisé plus haut, son obtention deviendrait une condition pour entrer au lycée ; à une visée certificative s'ajouterait donc une visée sélective), il paraît intéressant de détecter d'éventuelles évolutions sur un temps court. La comparaison des deux dernières éditions du brevet (avant 2016, épreuves de 2010) et présentement (à partir de 2017, épreuves de 2023) est éclairante.

Si le poids de la discipline « français » au DNB est sensiblement le même (16,6 % en 2010 vs 18,75 % en 2023, en incluant le contrôle continu), le poids de l'orthographe et des questions de grammaire est actuellement minoré par rapport à la compréhension et à l'interprétation du texte (questions de compréhension : 10 points sur 40 en 2010, soit 25 % de la note ; 32 points sur 100 en 2023, soit 32 %). Les questions de langue représentent 37,5 % de la note de français en 2010 (10 points sur 40 pour la dictée et un exercice de réécriture ; 5 sur 40 pour les questions grammaticales) et 28 % en 2023 (10 points sur 100 pour la dictée, 10 points pour l'exercice de réécriture ; 8 sur 100 pour les questions de « grammaire et compétences linguistiques »).

Le poids de la production d'écrit est légèrement réévalué avec la dernière mouture du DNB (+ 2,5 %) mais le flou est entretenu, comme par le passé, sur le poids de la langue dans l'évaluation de la production d'écrit, laissé à l'appréciation des jurys dans les différentes académies.

Si les lignes semblent avoir bougé au profit de la compréhension, principale compétence évaluée dans les évaluations internationales, on en reste en France à la compréhension écrite d'un texte littéraire (Colette en 2010 et George Sand en 2023). Ni la compréhension des textes informatifs, ni la compréhension orale ne sont évaluées ; or, comme les épreuves d'examen fournissent des modèles, elles définissent des attentes qui se substituent à celles exprimées par les programmes. On peut comparer à ce qui passe par exemple en Belgique pour le CE1D.

Les élèves belges passent le CE1D (certificat d'études du premier degré) en 8^e année (donc l'année de leurs 14 ans, à la fin de la 2^e année du secondaire, la scolarité primaire durant 6 ans). En 2010, le CE1D, évaluation externe certificative, se mettait en place. La participation des élèves était laissée à la décision des établissements. L'évaluation de la compréhension orale était alors envisagée ; elle a été introduite dès 2011. L'évolution des épreuves s'est clairement faite au profit de la compréhension (compréhension écrite : 41 points sur 90 ; compréhension orale : 14 points sur 90 ; soit +16,5 %), au détriment des « ressources liées à l'écriture », qui représentaient 26,1 % de la note (orthographe, syntaxe, structuration du texte, connecteurs), qui ne sont désormais plus évaluées. L'écriture d'un texte argumentatif de 150 mots est attendue et valorisée à hauteur de 35 points sur 90, soit 38,9 % de la note finale (le poids était de 35 sur 115 en 2010, soit 30,4 %), 11 points sur les 35 sont consacrés à la langue.

Tout comme le DNB, l'épreuve anticipée de français (EAF) que les lycéens français des voies générale et technologique passent à l'issue de la classe de 1^{re} est un exemple intéressant d'ajustement entre l'évolution de l'examen et l'évolution des pratiques d'enseignement. Cette épreuve est en quelque sorte le couronnement des deux années d'études de français au lycée. En classe de seconde et de première, le programme de français est entièrement organisé autour de la passation des épreuves de l'EAF : les seuls exercices prescrits/pratiqués sont ceux que les élèves auront à faire le jour de l'examen ou qui préparent à l'effectuation de ces exercices.

L'écriture d'invention a été introduite parmi les sujets possibles, à côté de la dissertation et du commentaire, en 2000-2001. Il a alors fallu beaucoup batailler pour empêcher qu'on voie en elle le signe d'une régression qui ramènerait les lycéens au niveau des collégiens (Daunay, 2003 ; Petitjean, 2003, 2005 ; Denizot, 2005 ; Reuter, 2005). Un peu moins de vingt ans plus tard, à la faveur d'une réforme du baccalauréat, l'épreuve anticipée de français a été retouchée et l'écrit d'invention en a fait les frais. Il disparaît alors non seulement de l'examen mais des exercices d'écriture pratiqués en classe, provoquant un changement d'équilibre dans la discipline français qui perd la possibilité de conférer aux élèves la posture auctoriale que cet exercice leur demandait d'adopter (Houdart-Merot et AM Petitjean, 2018 ; Plane, 2022). Un rapprochement peut être fait avec ce que constatait Arthur Applebee (2017) aux États-Unis : la mise en place d'une évaluation commune par le National Assessment of Educational Progress qui exigeait des élèves la production d'un paragraphe très calibré a conduit les enseignants du secondaire à faire pratiquer assidument ce type d'exercice, au détriment de la production de textes d'une plus grande ampleur, qui a fini par disparaître des tâches scolaires.

1.4. Les épreuves certificatives entre modèles, définitions institutionnelles et définitions empiriques

Les épreuves certificatives font toujours l'objet d'une définition minutieuse de la part de l'institution qui les promeut, à la fois pour guider leur préparation et assurer que la correction soit homogène, mais aussi pour prévenir les éventuels recours administratifs ou juridiques. Cependant, à côté de la définition officielle d'une épreuve, il se construit une définition empirique, mobilisée à la fois dans la préparation et dans l'évaluation de cette épreuve, et qui se fonde sur les modèles qu'on convoque ou sur la mémoire des sessions précédentes. Les rapports de jury de concours jouent un rôle particulièrement important, comme on peut le voir dans l'article de Caroline Atallah et Marie-Armelle Camussi-Ni, parce qu'ils orientent la préparation des concours et déterminent en partie les sessions qui vont suivre. Outre ces documents officiels, les supports d'exercice qui ont été utilisés, les questions qui ont été posées, les corrigés rédigés par les concepteurs des sujets constituent des archives qui encadrent la production de toute nouvelle

épreuve. Les productions attendues sont modélisées par l'ensemble sédimenté des prestations orales ou écrites, d'inégale qualité, réalisées par les candidats précédents.

Ce corpus implicite et les échanges qu'il suscite au sein de la communauté académique enrichissent la définition initiale de l'épreuve d'évaluation et font que celle-ci évolue au fur et à mesure que ce corpus s'enrichit. Le « Grand oral » du baccalauréat nous en offre un exemple.

Le Grand oral est un examen récent – la première session a eu lieu en 2021 – et les traces de sa genèse qui a vu s'affronter plusieurs modèles sont encore visibles. Cet examen est issu d'un rapport élaboré par Cyril Delhaye (2019), qui se présente comme professeur d'art oratoire à Sciences Po⁶, et il constitue une pièce importante de la réforme du baccalauréat dont le projet a été confié au directeur de l'IEP de Lille, Pierre Mathiot (2018). Sa création pourrait être considérée comme un prolongement des programmes de 2015 qui avaient donné une place notable à l'oral et à son enseignement au collège, mais le rapport Delhaye s'en éloigne en mettant l'accent sur la performance oratoire, avec l'idée que sa dimension corporelle permet d'échapper au déterminisme social : « Si l'apprentissage de l'oral est accessible au plus grand nombre, c'est parce qu'il se fonde sur la coordination corporelle. Parler en public est en premier lieu une activité physique » (p. 3). Cette idée rencontre d'autant plus de succès qu'elle paraît apporter une réponse simple aux problèmes complexes et récurrents d'inégalité des chances que connaît l'école. De plus, elle entre en écho avec un engouement médiatique et cinématographique pour l'éloquence : des films atteignent une large audience en racontant comment, à force d'efforts pour maîtriser l'art de la parole, le héros vainc le destin difficile que lui réservait son handicap (*Le discours d'un roi*, 2010) ou sa classe sociale (*Le brio*, 2017) ; les concours d'éloquence se multiplient, accompagnés d'émissions de télévision ou de films (*À voix haute*, 2016) qui donnent à voir et à entendre les prestations de jeunes orateurs, souvent issus de quartiers défavorisés, qui triomphent grâce à leur talent oratoire ; au théâtre, Richard Berry déclame des plaidoiries célèbres tandis qu'Éric Dupont-Moretti, qui n'est pas encore ministre, déclame les siennes... Ce contexte fait que le grand oral sera d'abord associé à l'idée de compétition et de prestation prestigieuse par des officines qui, sans avoir même lu le projet officiel, profitent de l'absence provisoire de formation dédiée à ce nouvel objet pour proposer leurs services sous forme de livres (l'avocat Bertrand Perrier en publie neuf entre 2017 et 2022) ou de cours, convoquant en renfort l'image de tribuns tel Martin Luther King soulevant l'enthousiasme d'une foule immense et subjuguée. Ce modèle qui se diffuse dans les médias entre en concurrence avec le projet présenté par Delhaye. Ils ont comme point commun l'importance accordée au corps : le projet prévoit pour les élèves uniquement une formation de sept heures portant sur la respiration, la voix et les neurones miroirs. La présence des neurones miroirs dans cette liste, alors que les implications qu'on leur prête suscitent des

⁶ « Sciences Po » est une dénomination familière pour « Institut d'études politiques » ou IEP.

réerves de la part des spécialistes (Decroix *et al.*, 2022), peut se lire comme une marque de révérence envers les neurosciences. L'idée de compétition véhiculée par les concours d'éloquence n'est pas totalement absente du projet de Grand oral du bac, dans la mesure où celui-ci a pour référence, comme son nom l'indique, des épreuves visant non seulement à la validation du cursus suivi par des étudiants de niveau avancé mais aussi à la sélection ou au classement de ces étudiants : Grand oral de l'ENA⁷ qui ouvre à la haute fonction publique, Grand oral du barreau qui ouvre à la carrière d'avocat, et surtout Grand oral de Sciences Po, où enseigne le promoteur du projet. S'ajoutent à ces modèles ceux d'examens étrangers de fin d'études secondaires et en particulier le *colloquio*, épreuve interdisciplinaire que passent en Italie les candidats à l'*esame di maturità*. Le Grand oral du baccalauréat emprunte à ces modèles l'obligation faite au candidat de réaliser sa prestation debout. Mais il impose également aux candidats une contrainte supplémentaire, celle de parler sans notes, comme un orateur aguerri, contrairement au *colloquio* ou à l'examen de maturité en Suisse qui prennent comme point d'appui un mémoire réalisé par le candidat. Le protocole prévoit un moment de dialogue avec le jury entre deux monologues dont le dernier est consacré à l'exposé des projets d'étude ou de carrière du candidat, comme dans les entretiens de recrutement. La question se pose donc de savoir quelle image idéale du lycéen sur le plan intellectuel et social a été convoquée pour élaborer ce type d'épreuve.

L'épreuve relève du domaine disciplinaire choisi par le candidat en tant qu'« enseignement de spécialité », mais le projet envisage une évaluation portant surtout sur la performance langagière. Il met d'ailleurs en garde contre « le risque que l'attention de l'évaluateur se focalise sur les seuls contenus, rendant ainsi l'épreuve orale redondante avec celle de l'écrit » (Delhaye, 2019, p. 4). Pourtant, dans les textes réglementaires, c'est au professeur disciplinaire qu'il incombe de former ses élèves à la prise de parole et de dispenser les apprentissages techniques nécessaires pour réaliser la performance attendue.

L'inspection générale a été amenée à revoir le projet et à rééquilibrer les critères d'évaluation avant de le rendre officiel. Les deux sessions d'épreuve ont permis de lui donner une définition plus nourrie, plus réaliste. Elles ont aussi conduit à une révision du dispositif : dans la dernière version des textes réglementaires, les élèves ont le droit de consulter les notes qu'ils ont élaborées durant la préparation, d'être assis pendant le dialogue avec le jury, et la dernière partie de l'épreuve a été supprimée. Le grand oral reste une épreuve exigeante, mais elle emprunte désormais, sans le dire, à des modèles scolaires dont on avait essayé de la couper, en particulier celui de l'oral du DNB, ou des TPE (travaux personnels

⁷ L'ENA (École nationale d'administration) est remplacée depuis 2022 par l'INSP (Institut national du service public) et la procédure de classement de sortie a été annulée par une publication du *Journal officiel* du 26 janvier 2023. À noter que le dernier rapport de jury de l'ENA contient cette mention étonnante à propos de l'épreuve orale : « les jurys élaborent leurs propres critères qui ne sont rendus que partiellement publics » (Chevrier, 2022, p. 14).

encadrés), supprimés en 2020 pour laisser la place au Grand oral, qui étaient pluridisciplinaires et avaient sur le Grand oral l'avantage de ne pas conférer de supériorité aux élèves qui ont acquis en dehors de l'école des codes de communication experts.

1.5. Les évaluations certificatives : un contrat entre les enseignants et leurs élèves

Les rôles que jouent les enseignants se déclinent de différentes façons selon l'organisation générale des évaluations certificatives, puisque, s'ils sont toujours mobilisés pour préparer leurs élèves à ces évaluations, ils ne sont pas nécessairement engagés dans la correction des épreuves et, s'ils le sont, ils peuvent être impliqués directement dans l'évaluation de leurs propres élèves ou au contraire en être écartés.

Ainsi en Belgique francophone, les épreuves du certificat d'enseignement secondaire supérieur se déroulent dans les établissements où les élèves sont scolarisés, sous la responsabilité de leurs enseignants, et le diplôme est délivré par le conseil de classe ; en Suisse, les candidats au certificat de maturité peuvent obtenir leur diplôme en passant les épreuves d'examen dans les établissements où ils sont scolarisés, si ceux-ci sont agréés, ou en se soumettant à un examen organisé par une commission ; au Québec, seule l'épreuve uniforme de français est commune à tous les établissements. En France, il est d'usage que les épreuves du baccalauréat soient évaluées par des enseignants que les élèves n'ont jamais eus comme professeurs et la rupture récente de cette tradition par l'introduction d'une part de contrôle continu n'est pas sans soulever des contestations. En effet, depuis la réforme de 2021, 40 % du total de points entrant dans la note finale du baccalauréat provient du contrôle continu et cette proportion a été élevée à 60 % à partir de la session 2024. Ces mesures ne procèdent pas tant de considérations d'ordre pédagogique – le contrôle continu serait réputé plus juste que l'examen, sujet à des aléas – mais relève plutôt de considérations pratiques : il est nécessaire que les notes permettant l'obtention du diplôme soient disponibles avant la fin de l'année scolaire, afin que le résultat final puisse être anticipé pour permettre le fonctionnement de Parcoursup, le dispositif d'accès à l'enseignement supérieur qui permet à chacun des 720000 futurs bacheliers de faire 10 vœux – assortis possiblement de 10 sous-vœux – parmi les 23000 formations d'enseignement supérieur proposées. La prise en compte des notes de contrôle continu a suscité des réactions hostiles, d'une part en raison de l'extrême complexité du premier dispositif d'épreuves communes de contrôle continu (appelées alors E3C, aujourd'hui EC) qui supposait que toutes les classes avancent d'un même pas, d'autre part, en raison du sentiment que les notes de contrôle continu n'offraient pas les mêmes garanties d'équité que l'examen, d'où une suspicion à l'encontre de l'enseignement privé considéré comme pouvant avoir la main sur les résultats de ses élèves, ce qui n'est pas le cas quand l'examen externalisé permet à lui seul d'obtenir le diplôme. Le français n'est concerné

par le contrôle continu qu'à travers l'enseignement de spécialité « humanité, littérature et philosophie », la discipline « français » proprement dite étant évaluée par l'épreuve anticipée de français, que nous avons évoquée plus haut.

Cette quasi-systématisation du recours à l'évaluation externe, comme c'est le cas en France pour les examens de fin de cycle, à savoir le baccalauréat et le DNB n'est sans doute qu'un des facteurs qui permettent aux évaluations certificatives d'exercer une pression sur l'enseignement. Mais l'examen a un avantage sur le contrôle continu, ce qu'avait perçu Felix Pécaut, le rédacteur de l'article « Examen » dans le dictionnaire de Ferdinand Buisson (1911) : il existe une sorte de pacte entre les élèves et leur enseignant, ou les étudiants et leur formateur, lorsqu'il s'agit de préparer les apprenants à passer une épreuve d'examen ou de concours. Enseignants et apprenants sont partenaires d'un même combat contre un adversaire abstrait, l'évaluateur, dont il faudra apprendre à deviner les attentes et déjouer les pièges, ce qui n'est pas le cas dans le contrôle continu, où l'enseignant est en même temps celui qui décide de la réussite.

2. Les évaluations sélectives : des lieux de confrontation

Les évaluations sélectives sont fondées sur la compétition puisqu'elles opposent des candidats entre eux. Lorsqu'elles servent à recruter des enseignants, elles ne font pas que mettre en œuvre des hiérarchies entre les postulants par le biais de leur classement, elles prennent également place sur un fond de conflits de hiérarchies entre les ordres d'enseignement et expriment des hiérarchies entre les savoirs ou compétences jugées nécessaires pour enseigner à travers le choix des épreuves et leurs contenus.

2.1. Un instrument de politique éducative

Le recours aux concours pour recruter les enseignants est une longue tradition française, qui s'impose aux enseignants du public comme à ceux du privé, même si les premiers n'ont pas le choix de leur affectation alors que les seconds postulent à leur convenance sur les postes qui les intéressent, comme cela est le cas dans les pays qui recrutent sur titre et non sur concours : au Québec le brevet d'enseignement, en Belgique l'AESI (agrégation de l'enseignement secondaire inférieur) et l'AESS (agrégation de l'enseignement secondaire supérieur), en Suisse le diplôme d'enseignement confèrent l'autorisation d'enseigner et permettent au candidat de postuler directement sur un emploi.

L'évolution des concours en France et leur multiplication reflètent les débats dont la profession enseignante a fait – et continue de faire – l'objet, comme le montrent la tension entre des tentatives d'unification de la profession enseignante et la résistance à ces tentatives ou encore la mise en opposition des savoirs professionnels et des savoirs académiques. À l'agrégation, dont le concours existe depuis 1821, se sont ajoutés plusieurs concours pour recruter des enseignants du secondaire afin de répondre à la

pénurie de professeurs tout en maintenant les exigences spécifiques de l'agrégation, avec d'abord, brièvement, sous le régime de Vichy, le CAEC (certificat d'aptitude à l'enseignement en collège), puis en 1950 le CAPES (certificat d'aptitude pédagogique à l'enseignement secondaire) presque immédiatement réformé car il débutait par des épreuves pédagogiques, lesquelles ont été considérées comme faisant ombrage aux épreuves scientifiques (Chervel, 1993 ; Chapoulie, 2005 ; Verneuil, 2016). Le CAPES prend donc en 1952 une forme qui le rapproche de l'agrégation, jusqu'à ce que soit introduite, en 1992 une épreuve orale d'abord appelée « épreuve professionnelle » qui prendra différentes dénominations. Il se diversifie en fonction des secteurs d'enseignement (enseignement technique, lycée professionnel, etc.) et admet des variantes avec, par exemple, la création en 2014 d'une option FLES (français langue étrangère ou seconde) du CAPES de lettres modernes (Spaëth, 2015). Les tentatives de rapprochement des ordres d'enseignement initiées par Jean Zay sous le Front populaire se sont poursuivies dans l'après-guerre avec des réformes unifiant le cursus des élèves et des mesures visant à harmoniser les formations et les carrières des enseignants. Cette harmonisation s'est concrétisée avec la loi d'orientation de 1989 qui a instauré pour le nouveau corps de professeurs des écoles qui succède à celui d'instituteurs des modalités de recrutement semblables à celles des enseignants du secondaire. La volonté d'homogénéiser formellement tous les concours s'est traduite plus récemment à la fois par la réduction du nombre d'épreuves et par l'introduction de nouveaux objets d'évaluation transversaux, mettant ainsi en concurrence les différents savoirs nécessaires au métier.

2.2. Un révélateur des divergences entre différentes conceptions des savoirs nécessaires pour enseigner

En France, la formation et les concours pour enseigner dans le primaire, après avoir été immuables pendant des décennies, ont connu un changement de grande ampleur avec la loi d'orientation de 1989 qui demande le même niveau d'étude pour passer les concours du primaire et du secondaire, à savoir la licence, et crée le titre de professeur des écoles. Cette loi instaure un concours qui exige des postulants une réflexion didactique, notamment avec l'épreuve de français qui comporte une synthèse de textes relatifs à l'enseignement du français et une analyse de texte d'élève. Mais après une période de stabilité, les réformes se multiplient et la didactique en fait souvent les frais, sur un fond de concurrence artificielle entre savoirs universitaires et savoirs professionnels, entre « grande littérature » et littérature de jeunesse, entre maîtrise de la langue et savoirs pour enseigner la langue.

L'histoire du concours de recrutement des professeurs des écoles (CRPE) devient vraiment tourmentée à partir de 2010, date à laquelle le master est devenu obligatoire pour qui veut enseigner, dans le primaire comme dans le secondaire. Le stage et la formation que les lauréats des concours recevaient en tant que stagiaires sont supprimés et ne seront rétablis que par la loi de 2013 sur la refondation de l'école.

L'admissibilité continue de comporter une épreuve écrite de français et une autre de mathématiques, la représentation de la polyvalence prenant des formes variées soit à l'écrit, soit à l'oral.

Avec la réforme de la maîtrise en 2010, l'admissibilité se fait désormais en M1 et l'admission en M2. Le nombre des épreuves d'admissibilité passe de trois à deux : le français est regroupé avec l'histoire, la géographie et l'instruction civique et morale dans une première épreuve (note sur 20, dont 12 points pour le français). La seconde épreuve d'admissibilité regroupe les mathématiques, les sciences expérimentales et la technologie (note sur 20, dont 12 points pour les mathématiques). Le français comme les mathématiques font l'objet d'une leçon lors des épreuves d'admission.

À partir de 2014, nouvelle réforme : l'admissibilité et l'admission ont lieu en fin de M1. Deux épreuves suffisent désormais à l'admissibilité : une en français et une en mathématiques, et les apprentissages dits fondamentaux disparaissent des épreuves orales au profit d'« une mise en situation professionnelle dans un domaine au choix du candidat ». L'épreuve de français est notée sur 40 points : 11 pour la première partie (question relative aux textes proposés), 11 pour la deuxième (connaissance de la langue) et 13 pour la troisième (analyse de supports d'enseignement) ; 5 points sont dévolus à la correction syntaxique et à la qualité écrite de la production du candidat. Une note globale égale ou inférieure à 10 est éliminatoire. La production de texte, notée sur 11 points, a le même poids que la langue. La troisième question est une épreuve de didactique, réintroduite après sa suppression en 2010 (analyse de pages de manuel, de textes d'élèves, de transcription d'échanges en classe ou de préparation d'une séance ou d'une séquence).

Depuis 2022, admissibilité et admission se déroulent en fin de M2. Une 3^e épreuve écrite, dite d'application, a été introduite. Son objectif est d'apprécier la capacité du candidat à proposer une démarche d'apprentissage progressive et cohérente. Le candidat a le choix au début de l'épreuve entre trois sujets portant respectivement sur l'un des domaines suivants : sciences et technologie ; histoire, géographie, enseignement moral et civique ; arts.

La nouvelle épreuve de français « prend appui sur un texte (extrait de roman, de nouvelle, de littérature d'idées, d'essai, etc.) d'environ 400 à 600 mots [et] comporte trois parties :

- une partie consacrée à l'étude de la langue, permettant de vérifier les connaissances syntaxiques, grammaticales et orthographiques du candidat [sur 6] ;
- une partie consacrée au lexique et à la compréhension lexicale, reliée au sens du texte [sur 4] ;
- une partie consacrée à une réflexion suscitée par le texte à partir d'une question posée sur celui-ci et dont la réponse prend la forme d'un développement présentant un raisonnement rédigé et structuré [sur 10] » (MENJ, 2024).

Depuis 2022, l'épreuve écrite de didactique a disparu. Le texte demandé aux candidats est produit à partir d'un seul texte soumis à leur réflexion. Les questions de langue sont sensiblement les mêmes que précédemment, de 2014 à 2021 (emplois de temps verbaux, repérage de subordonnées, justification de

l'orthographe de participes passés, identification de classes grammaticales) sauf que le lexique fait l'objet de questions clairement fléchées, ce qui permet de garder trois questions.

L'observation des trois réformes conduites en 12 ans révèle les points de tension que sont :

- l'évaluation des autres disciplines que le français et les mathématiques, qui font la polyvalence du professeur des écoles ;
- le risque que l'évaluation ne sollicite pas une formation suffisante des futurs professeurs des écoles ;
- la façon d'évaluer les compétences professionnelles avec la question de didactique à l'écrit ou la leçon à l'oral.

À l'heure où nous écrivons ce texte, une nouvelle réforme est en train de voir le jour, qui doit entrer en vigueur en 2025 : le concours se déroulera en fin de licence (3^e année) et ce déplacement devrait affecter à nouveau les épreuves.

3. Les évaluations à fonction informative : du pilotage par l'évaluation au contrôle par les évaluations

Le nombre et l'ambiguïté des évaluations institutionnelles à visée informative méritent qu'on s'y intéresse particulièrement (Bart, 2015 ; Bart et Daunay, 2015, 2016 ; De Ketele, 2009) et qu'on tâche d'apprécier leurs fonctions, qui vont bien au-delà du simple apport d'informations. En effet, des enquêtes internationales, dont le Programme international pour le suivi des acquis (PISA) est le plus connu, et des tests nationaux, comme ceux que passent en France les élèves à leur entrée à l'école élémentaire, servent également à piloter le système éducatif, voire à gouverner les enseignants.

3.1. Le souci de contrôler l'efficacité de la dépense publique

Le développement des épreuves standardisées est à mettre en rapport avec le souci de contrôler l'efficacité de la dépense publique au moyen d'une politique dite de rationalisation des choix budgétaires (RCB) à l'instar du *Planning, Programming and Budgeting System* instauré aux USA (Monnier, 1992) qui exerce une influence de plus en plus pressante à partir du choc pétrolier de 1973. La doctrine du *New Public Management*, que Philippe Bezes (2005) décrit comme un ensemble hétérogène de théories économiques et gestionnaires, renforce dans les années 1980 cette tendance en poussant vers un « amincissement de l'État », compensé par le transfert de la responsabilité des réformes adoptées vers les agents qui les mettent en œuvre, assorti d'un contrôle plus serré des agents en question. L'enseignement constituant un gros poste de la dépense publique, il est bien évidemment concerné au premier chef par cette politique.

En France, pour apprécier l'efficacité de l'enseignement, en tant qu'elle renseigne sur l'efficacité de la dépense afférente, la Direction de l'évaluation et de la prospective (DEP ; aujourd'hui : DEPP, le second P

renvoyant à *performance*) a constitué des panels et procédé à des suivis d'élèves dès 1973 et à des évaluations à des moments-clés du parcours à partir de 1979. Cette période est celle où s'installe le paradigme de la RCT (Randomized Control Trial) qui exige que l'efficacité des expérimentations soit vérifiée par le recours à différentes mesures sous la forme de pré-tests et post-tests, lesquels fourniront des modèles d'épreuves pouvant être utilisées par les instances chargées de l'évaluation nationale (Céard *et al.*, 2003 ; Rémond, 2005). S'ajoute à ce souci de contrôle de l'efficacité, la nécessité de disposer d'éléments objectifs pour répondre à la crainte d'une baisse de niveau engendrée par la massification (Thélot, 1994, 2008 ; Trosseille et Rocher, 2015).

L'imperfection des méthodologies employées jusqu'alors qui ne permettaient pas de procéder à des comparaisons facilite l'adoption en France d'un projet consistant à soumettre non plus un panel, mais l'ensemble des élèves de CE2 et de 6^e à des évaluations standardisées dans les années qui suivent la réforme de l'enseignement et de la formation instaurée par la loi d'orientation de 1989. Le fait que ces évaluations prenaient en compte tous les élèves et non pas seulement un panel a permis d'en faire des instruments d'information directe à l'intention des enseignants et des parents, au moment où se développait une culture de l'évaluation nourrie par les travaux de recherche en didactique. Ce dispositif a servi de modèle à toutes les évaluations nationales qui ont été mises en place depuis lors, même si les modes de conception des épreuves ont évolué.

Les évaluations nationales sont complétées par des évaluations internationales censées apporter des points de comparaison et qui procèdent le plus souvent d'institutions ayant une approche économique de l'éducation. Dans ce cadre le mot « évaluation », employé dans l'enseignement pour référer à une prise d'information dans la classe, rejoint le sens que ce même terme a dans le vocabulaire économique où il renvoie à l'appréciation du coût des objets ou des services. L'exemple le plus emblématique est celui de PISA, dont Daniel Bart et Bertrand Daunay proposent une analyse critique dans ce numéro. Menée par l'OCDE (Organisation de coopération et de développement économique), cette enquête apporte des informations qui sont le plus souvent traitées par les instances politiques sous l'angle de la concurrence et de la compétition, comme s'il s'agissait de comparer des performances économiques. La conversion des résultats en données numériques et les classements y contribuent largement. Si l'on adopte cette approche restrictive, on perd un aspect important de cette évaluation qui est d'inciter à une réflexion sur les standards. Pour le dire autrement, plus que les résultats numériques eux-mêmes, ce sont les questions que pose ce type d'évaluation qui peuvent aider à faire progresser l'enseignement. Jean-Marie de Ketele (2009) note ainsi que ce sont les enquêtes PISA qui ont rendu manifeste la nécessité d'harmoniser les systèmes éducatifs en Suisse, ce qui s'est concrétisé par le concordat intercantonal HarmoS dont la mise en place entamée en 2001 n'a été pleine et entière qu'en 2015.

Il est à signaler que les interprétations des données fournies par les évaluations nationales ou internationales font souvent appel à un raccourci impensé pour défendre, promouvoir ou invalider des décisions en matière de politique éducative : ce sont les savoirs et compétences des élèves qui ont été évalués, mais on fait comme si c'était l'enseignement qui leur a été dispensé qui avait été évalué, alors qu'en réalité les choses ne sont pas si simples.

3.2. Des évaluations à visée informative pour instaurer de nouvelles pratiques évaluatives chez les enseignants

Les évaluations institutionnelles à visée informative ont pour vocation première de fournir des données aux décideurs institutionnels. Mais elles peuvent aussi avoir une fonction formative si les enseignants qui y participent ne sont pas considérés uniquement comme des agents qui exécutent un protocole, mais sont au contraire engagés dans une réflexion sur ce qu'apprend l'évaluation et sur ses fonctionnements. Autant dire que les grandes évaluations internationales ne peuvent pas remplir ce rôle. En revanche, la mise en place des premières évaluations CE2-6^e en France dans les années 1990 avait pour objectifs de non seulement collecter des données de grande ampleur mais aussi de développer une culture de l'évaluation (Thélot, 2008). Les premières sessions ont été contemporaines des travaux du groupe EVA (1991, 1996) qui s'intéressaient à décompacter l'évaluation de la production d'écrit, à faire réfléchir à des critères, à promouvoir le brouillonage et la réécriture. Ces pratiques et ces préoccupations ont pu être diffusées par les protocoles de l'évaluation nationale qui comportaient alors une production d'écrit. Malheureusement, cette partie de l'épreuve a été par la suite effacée des dispositifs d'évaluation, ce que déplore Kathy Similowski dans l'article qu'elle lui consacre.

La mise en place des évaluations a été accompagnée par un plan de formation de grande ampleur qui visait à aider les enseignants à s'approprier non seulement les résultats des évaluations pour avoir une appréciation plus fine des compétences de leurs élèves, mais aussi à en comprendre les principes pour être en mesure d'élaborer pour leurs propres besoins de nouveaux outils d'évaluation. Il s'agissait aussi de les faire passer d'une approche purement normative et sommative de l'évaluation – note et évaluation étant alors communément considérées comme des équivalents – à la pratique de l'évaluation formative.

La démarche s'est amplifiée par la suite, et d'autres dispositions ont été prises pour engager les enseignants à ce type d'approche de l'évaluation, non seulement à l'occasion des exercices eux-mêmes, mais au moment de porter des appréciations sur les élèves, notamment dans une optique de normalisation de ces appréciations à des fins d'équité. C'est ainsi qu'a été développé le LSU (livret scolaire unique), utilisé désormais dans tous les établissements scolaires du CP à la 3^e qui vise à l'harmonisation des pratiques d'évaluations. Mais cette visée qu'on pourrait qualifier d'universaliste puisque le LSU doit

pouvoir être adapté à toutes les situations, a pour revers sa complexité (le guide élaboré par l'Inspection pédagogique régionale de Poitiers pour la prise en main du LSU fait 29 pages).

3.3. Des évaluations à visée informative pour guider ou pour contrôler l'action des enseignants ?

Au cours des dernières décennies, le nombre et la fréquence de ces évaluations se sont accélérés en France si bien que des niveaux d'enseignement qui jusqu'alors échappaient à ce type d'évaluation sont désormais concernés, comme en témoigne le test de positionnement en classe de seconde (Andreu *et al.*, 2022). Le développement de ce type d'évaluation institutionnelle est loin d'être une spécificité française : les programmes internationaux PIRLS (Progress in International Reading Literacy Study), l'enquête Reading Literacy de PISA et les évaluations ministérielles mises en place dans différents pays de la francophonie le prouvent (Perrin-Doucey, 2020 ; Brissaud et Lefrançois, 2014 ; Lefrançois et Brissaud, 2015). Mais plusieurs raisons expliquent une « frénésie française des évaluations » (Gentaz, 2022). Ces évaluations ont d'abord pu trouver un terreau favorable en France parce qu'elles entrent en concordance avec la tradition républicaine française qui attache du prix au caractère national de l'évaluation des connaissances et compétences au moyen d'examens communs à tous les impétrants, contrairement à d'autres pays, comme les États-Unis où le High School Graduation s'obtient au moyen du contrôle continu. De même en France, l'accès à un grade ou un corps de fonctionnaire se fait par voie de concours⁸, ce qui explique d'ailleurs que le Master MEEF (Métiers de l'enseignement, de l'éducation et de la formation) ne suffise pas à permettre la titularisation de ses détenteurs qui doivent également passer un concours. Mais le développement récent et envahissant des évaluations en France ces dernières années est aussi à mettre en rapport avec une volonté de piloter de façon très serrée l'action des enseignants, surtout dans l'enseignement primaire. Les évaluations qui scandent la scolarité des élèves, parfois de façon très rapprochée, ont un pouvoir contraignant bien plus puissant que celui des programmes puisqu'elles imposent aux enseignants de préparer leurs élèves à des épreuves calibrées qui interviennent à un moment dont ils n'ont pas la maîtrise. L'extrême fréquence des évaluations à l'école primaire est rendue possible par le poids qu'exerce le Conseil scientifique de l'éducation nationale (CSEN), mis en place par Jean-Michel Blanquer lorsqu'il était ministre de l'Éducation. Cette instance a en effet la possibilité de commanditer des protocoles d'évaluation à la DEPP et d'en imposer la passation à la France entière, alors qu'il existe par ailleurs des enquêtes sur panel qui apportent des informations d'un très haut niveau de fiabilité, par exemple l'enquête menée par la DEPP sur les compétences numériques des élèves de fin de

⁸ Le principe du concours comme condition d'accès au statut de fonctionnaire a été défini par les articles 26 et 27 de la loi du 19 octobre 1946, qui mettait ainsi fin au fonctionnement ségrégationniste imposé par le régime de Vichy (Waline, 1957). L'article 13 de la loi Le Pors du 13 juillet 1983 reprend ce principe.

collège (M'Bafoumou *et al.*, 2023) qui a montré que les deux tiers des élèves avaient une bonne maîtrise et que ceux qui avaient un bon niveau en français à l'entrée au collège ont une meilleure maîtrise du numérique que les autres. Les évaluations imposées à l'ensemble des élèves d'un niveau donné, sur demande du CSEN, non seulement pilotent l'action des enseignants mais fournissent également des données aux chercheurs, qui ont d'ailleurs tendance à considérer que si les résultats recueillis ne confirment pas leurs hypothèses, c'est parce que les enseignants n'ont pas appliqué suffisamment bien leurs préconisations, aussi estiment-ils qu'il faudrait les y contraindre par des évaluations de plus en plus nombreuses et serrées.

Le développement de la frénésie évaluative en France a commencé par une première salve de trois évaluations standardisées, sur support papier, initiées en 2018, portant sur la maîtrise des correspondances graphème-phonème et des compétences reliées ainsi que la maîtrise de la compréhension du langage oral, en début de CP (première année de l'école élémentaire), en milieu de CP et en début de CE1, autrement dit tous les semestres. Depuis 2022 s'y est ajoutée une évaluation en début de sixième, première classe du secondaire. Elle porte sur la compréhension de l'écrit, la compréhension de l'oral, l'étude de la langue, est réalisée sur support numérique avec correction automatique et comporte un test de fluence qu'analyse Erika Godde dans ce numéro de *Repères*. Depuis 2023, un test portant sur les mêmes catégories de compétences se passe désormais l'avant-dernière année du primaire (CM1) et un autre la troisième année du secondaire (4^e). L'intérêt croissant pour l'évaluation de la compréhension est louable mais, comme le montrent les articles de Yann Mercier-Brunel, de Marie-France Bishop et Thomas Picard et de Patricia Schillings et de ses collègues de l'université de Liège, cette évaluation soulève beaucoup de questions.

S'ajoutent à ces évaluations, depuis 2021, les tests de positionnement en classe de seconde, c'est-à-dire à l'entrée au lycée, ou en classe de CAP, qui portent sur la littératie et la numératie.

N'oublions pas que les évaluations nationales sont complétées par les informations qu'apportent le PISA (élèves de 15 ans), le PIRLS (Programme international de recherche en lecture scolaire, classe de CM1) et l'ICILS (International Computer and Information Literacy Study, classe de 4^e).

Comparons avec les « épreuves ministérielles » du primaire et du secondaire au Québec, qui comportent des distinctions en fonction du statut de la langue française. Les élèves passent en 4^e primaire en fin d'année, une épreuve de lecture et écriture portant sur le français langue d'enseignement, avec comme matériel autorisé le dictionnaire et un tableau de conjugaison ; et en 6^e primaire une épreuve de français langue d'enseignement associée à d'autres disciplines, qui porte sur la lecture et l'écriture. L'écriture se passe en deux temps : le jour 1 est consacré à la planification et à la rédaction du brouillon, le jour 2 à la révision, la mise au propre et à la relecture ; toujours avec les outils.

On ne peut s'empêcher de rapprocher la surveillance et la défiance dont font l'objet en France les enseignants et les élèves à travers la multiplication des évaluations institutionnelles à des fins d'information, des projets de labellisation des manuels scolaires et de réfection des programmes qui prévoient de détailler par le menu toutes les actions que devront accomplir les enseignants, selon un calendrier qui ne leur laisse aucune marge de liberté⁹.

4. L'analyse des évaluations institutionnelles pour éclairer des problématiques de l'enseignement du français

Dans les articles que réunit ce numéro les évaluations institutionnelles sont examinées du point de vue de leurs enjeux didactiques et rapportées à leurs contextes de façon à dégager leurs effets sur l'enseignement et, partant, sur les apprentissages (De Pietro *et al.*, 2016 ; Mottier Lopez *et al.*, 2016). Ces articles apportent ainsi chacun une réponse à la question que nous posons plus haut « à quoi servent les évaluations institutionnelles ? », et que nous avons dédoublée en demandant également « à quoi fait-on servir les évaluations institutionnelles ? »

Les évaluations peuvent, en principe, s'intéresser à tous les objets d'enseignement du français. Mais, en fonction de leurs propriétés, les objets d'enseignement se prêtent plus ou moins bien à tel ou tel type d'évaluation. D'où la différence entre ce qu'évaluent les évaluations internationales, les évaluations certificatives, les évaluations nationales standardisées. Ces différences sont exemplifiées par un premier ensemble de contributions qui portent chacune sur un type particulier d'évaluation : l'article de Daniel Bart et Bertrand Daunay met en lumière les stéréotypes que promeut l'évaluation PISA, celui de Caroline Attala et Marie-Armelle Camussi-Ni passe au crible l'épreuve de grammaire du concours de recrutement des professeurs de français, et celui de Kathy Similowski étudie l'évolution d'un test standardisé de production d'écrit.

Daniel Bart et Bertrand Daunay mettent en question la visée informative et scientifique du programme PISA « qui s'est imposé comme une référence centrale dans le domaine des évaluations institutionnelles scolaires », mais qui fait aussi l'objet de débats qu'ils rappellent. Leur analyse d'un corpus de textes produits par ce programme et autour de ce programme depuis le premier cycle d'évaluation en 2000 jusqu'à l'édition 2015 fait apparaître des décalages entre le discours affiché « sur le mode de l'évidence » et les commentaires parfois faits sur les résultats. L'article s'attache à démontrer les contradictions entre les prétentions du programme, à la fois sur le plan méthodologique et sur celui de ses finalités, et les biais auxquels il n'échappe pas. Au bout du compte, il apparaît que les nombreux écrits produits par le

⁹ Lettre de saisine ministérielle adressée au Conseil supérieur des programmes le 8 janvier 2024.

programme, sans doute insuffisamment contrôlés à la lumière des non moins nombreux écrits critiques produits, participent à l'entretien de stéréotypes, par exemple celui de genre, que le programme semble non seulement mettre en cause mais affiche comme devant être combattus.

Caroline Attala et Marie-Armelle Camussi-Ni s'interrogent sur l'impact des modes de recrutement des enseignants sur la formation des enseignants de lettres et sur la formation en langue des élèves du secondaire. Elles analysent ce qui est dit de la question de grammaire dans un corpus constitué de rapports de jury du CAPES et de l'agrégation de lettres produits de 2008 à 2023 pour comprendre les attentes institutionnelles en matière de connaissances linguistiques et d'enseignement de la langue. La confrontation des préconisations extraites de ces rapports et de celles extraites d'articles de recherche en didactique de la langue montre un décalage certain entre les unes et les autres. Pour les deux chercheuses, les démarches de manipulation présentes dans les programmes, qui aident à la conceptualisation par les élèves du fonctionnement de la langue, sont encore insuffisamment présentes dans les concours. L'étude des maquettes de formation en licence de lettres de l'année universitaire 2022-2023 et des retours d'un questionnaire adressé aux enseignants-chercheurs de sciences du langage, révèle des situations très variées d'une université à l'autre. La linguistique est cependant globalement peu présente dans le cursus universitaire de licence et de master de lettres modernes. Dans ces conditions, les auteures posent la question des attentes institutionnelles en matière d'enseignement de la langue et du revirement qui devrait s'opérer dans une période difficile de recrutement des enseignants de lettres. Elles appellent de leurs vœux un renouvellement de la formation des futurs enseignants avec une véritable formation linguistique susceptible de renouveler l'intérêt des élèves pour la langue et son fonctionnement.

Kathy Similowski s'interroge sur la disparition de l'écriture de texte des évaluations nationales en France alors que les programmes sont exigeants en la matière et que l'institution elle-même pointe le temps insuffisant consacré à la production écrite à l'école. Afin de comprendre cette disparition, elle analyse les évaluations diagnostiques de CE2 et de 6^e (8-10 ans) de 1989 à 2007, leurs objectifs affichés et les compétences visées. Si les épreuves sont relativement stables, les compétences évaluées acquises à l'école sont soumises à fluctuation et manquent parfois d'explicitation pour les élèves eux-mêmes. L'auteure met par exemple en évidence la grande variabilité dans les attentes relatives à la syntaxe (moins pour la cohérence du texte), des fluctuations dans le nombre d'erreurs tolérées ou dans le nombre de lignes attendues, etc. Kathy Similowski regrette donc le peu de prise en compte des compétences rédactionnelles dans les évaluations nationales en France et souligne que ces épreuves ne sont pas vraiment à même d'aider les enseignants à enseigner la production de texte. Or la demande d'aide des enseignants est forte et la compétence à produire des textes de qualité constitue un atout pour la réussite scolaire.

Comme le signale Kathy Similowski, les évaluations institutionnelles ont eu tendance à se désintéresser de la production d'écrit. En revanche, elles ont été nombreuses à se focaliser sur la lecture. Aussi une partie

importante du dossier est-elle consacrée à ce thème, qui est exploré sous l'angle du test de fluence dont Erika Godde fait une analyse critique et sous celui de la compréhension qui fait l'objet de trois contributions, celle de Marie-France Bishop et Thomas Picard qui comparent trois protocoles utilisés au cycle 3, celle de Yann Mercier-Brunel qui s'interroge sur la pertinence des informations qu'apporte le test standardisé passé par les élèves de 6^e, et celui de Patricia Schillings, Charlotte Dejaegher, Olivier Leyh et Marine André, qui posent la même question mais à partir de l'enquête PIRLS.

Erika Godde remet en cause l'évaluation de la fluence communément adoptée dans les évaluations nationales en France (le nombre de mots correctement lus en une minute ou NMCLM). Cette mesure, qui dépend beaucoup du texte lu, néglige le versant prosodique de la fluence, notamment le phrasé, et semble ne pas être un bon indicateur de la compréhension. Erika Godde propose un mode d'évaluation en accord avec une définition plus large de la fluence, incluant la prosodie et la compréhension. L'outil préconisé, une échelle multidimensionnelle de fluence (EMDF), permet l'évaluation de toutes les dimensions de la fluence indépendamment les unes des autres et constitue « un indice fiable de la compréhension », utile pour un repérage des élèves « à risque ». L'EMDF a été validée avec deux textes narratifs, proposés à 295 enfants du CE1 à la 5^e ayant le français pour langue maternelle, fournissant deux textes étalonnés, destinés à l'usage des enseignants.

Marie-France Bishop et Thomas Picard proposent de comparer trois évaluations institutionnelles portant sur la compréhension au cycle 3 : deux pilotées par la DEPP, au CM1 et en 6^e, ainsi qu'une évaluation propre à l'académie de Versailles, organisée en CM2. Ces trois évaluations ont en commun l'objectif affiché d'aider les enseignants à « mieux répondre aux besoins de leurs élèves » et « d'accompagner le pilotage pédagogique dans les établissements et au niveau national ». Les auteurs font l'hypothèse que les conceptions sous-jacentes aux choix des items sont différentes et se proposent de pointer ces différences. Leur analyse des items des épreuves de compréhension, orales comme écrites, qui repose sur le modèle de Walter Kintsch (1998), permet de conclure que les concepteurs des différentes évaluations « n'accordent pas la même valeur aux différentes compétences de compréhension ». C'est ainsi que la part des questions concernant les phénomènes de haut niveau varie d'un groupe de concepteurs à l'autre et apparaît bien faible dans les évaluations de 6^e. Dans la mesure où, comme le rappellent les auteurs, ce qui s'évalue devient ce qui s'enseigne, la question de la cohérence des évaluations proposées aux mêmes élèves de cycle 3 est posée ainsi que celle de leur adéquation aux attendus des évaluations internationales.

Yann Mercier-Brunel se focalise sur l'une des évaluations examinées par Marie-France Bishop et Thomas Picard : les exercices de compréhension du texte littéraire écrit proposé en version numérique aux élèves de 6^e en 2022. À partir de tous les documents fournis par la DEPP, il se concentre sur une analyse détaillée fine des 10 items du test de compréhension d'un texte littéraire pour déterminer la « logique évaluative du test » et évaluer l'aide apportée aux enseignants. Sans remettre en cause les choix méthodologiques

de la DEPP, il questionne ainsi lui aussi les choix opérés par les concepteurs des tests. Il s'appuie sur deux modèles issus des sciences de l'éducation, celui de Figari et celui de De Ketele, pour questionner la validité et la pertinence de certains items, et la fiabilité d'un test qui ne met pas en relation les réponses d'un élève les unes avec les autres et ne permet donc pas de définir des profils.

L'article de Patricia Schillings, Charlotte Dejaegher, Olivier Leyh et Marine André s'intéresse aux performances en lecture des élèves telles qu'évaluées lors des cycles PIRLS, et notamment à leurs compétences d'interprétation. Après avoir rappelé que le terme *interpréter* peut renvoyer à des processus différents selon les cadres théoriques envisagés (celui de PIRLS est distinct de celui de la lecture littéraire), les auteurs proposent une observation fine des réponses produites par un sous-échantillon d'élèves belges recueillies lors des cycles PIRLS de 2016 et 2021 pour les deux textes rendus publics correspondant à l'objectif *Lire pour l'expérience littéraire* (il en est un autre : *Lire pour acquérir et utiliser de l'information*). Dans une approche psychodidactique des évaluations (Vantourout et Goasdoué, 2014), ils se livrent à une analyse fine qualitative des réponses des élèves en inférant les démarches de compréhension mises en œuvre pour répondre et mettent en évidence, derrière l'erreur, des formes de compréhension de l'élève et des compétences en développement. Leur analyse des réponses ouvertes fournies par les élèves aboutit à une typologie d'erreurs susceptible d'éclairer les enseignants sur l'activité de réponse des élèves. Les auteurs espèrent avoir ainsi déplacé « le curseur de cette enquête internationale du pilotage vers la régulation des apprentissages. »

Bibliographie

- ALLAL, L. et MOTTIER LOPEZ, L. (dir.). (2007). *Régulation des apprentissages en situation scolaire et en formation*. De Boeck Supérieur.
- ANDREU, S., BRET, A., DURAND DE MONESTROL, H., FERNANDEZ, A., HEIDMANN, L., MARIN, N., PHILIPPE, C., ROCHER, T., SALLES, F. et VOURC'H, R. (2022). Test de positionnement de début de seconde 2021 : des résultats contrastés selon les caractéristiques des élèves et des établissements. *Note d'information de la DEPP*, 22.15. Récupéré sur : <<https://doi.org/10.48464/ni-22-15>>.
- APPLEBEE, A. N. (2017). What shapes school writing? Examining influences on school writing tasks over time in U.S. secondary school. Dans S. Plane, C. Bazerman, F. Rondelli, C. Donahue., A. N. Applebee, C. Boré, P. Carlino, M. Marquilló Larruy, P. Rogers et D. Russell (dir), *Research on Writing: Multiple Perspectives* (p.13-24). WAC Clearinghouse. Récupéré sur : <<https://wac.colostate.edu/books/wrab2014/>>.
- AUDIGIER, F., CRAHAY, M. et DOLZ, J. (dir.). (2006). *Curriculum, enseignement et pilotage*. De Boeck.
- BART, D. (2015). Les évaluations nationales : un « outil » insaisissable ? Une analyse de la documentation institutionnelle sur le dispositif. *Recherches*, 63, 51-72. Récupéré sur : <https://revue-recherches.fr/wp-content/uploads/2018/01/051-072_R63_Bart.pdf>.
- BART, D. et DAUNAY, B. (2016). L'évaluation de la lecture d'un texte littéraire selon le programme PISA : interrogations didactiques. Dans E. Falardeau, J. Dolz, J.-L. Dumortier et P. Lefrançois (dir),

- L'évaluation en classe de français, outil didactique et politique* (p. 177-195). Presses universitaires de Namur.
- BART, D et DAUNAY, B. (2015). Un plaisir à livre ouvert ? L'évaluation du plaisir de lire dans le PISA. *Recherches*, 63, 147-161. Récupéré sur : <https://www.revue-recherches.fr/wp-content/uploads/2018/01/147-161_R63_BartDaunay.pdf>.
- BELHOSTE, B. (1990). L'enseignement secondaire français et les sciences au début du XX^e siècle. La réforme de 1902 des plans d'études et des programmes. *Revue d'histoire des sciences*, 43(4), 371-400. Récupéré sur : <<https://doi.org/10.3406/rhs.1990.4502>>.
- BEZES, P. (2005). Le renouveau du contrôle des bureaucraties. L'impact du New Public Management. *Informations sociales*, 126(6), 26-37. Récupéré sur : <<https://doi.org/10.3917/inso.126.0026>>.
- BRISSAUD, C. et LEFRANÇOIS, P. (2014). Évaluations ministérielles de l'écriture et de la langue dans la francophonie en fin de scolarité primaire : quelles finalités ? *Langage et société*, 148(2), 107-124. Récupéré sur : <<https://doi.org/10.3917/lis.148.0107>>.
- CÉARD, M.-T., RÉMOND, M., et VARIER, M. (2003). *L'appréciation des compétences des élèves et des jeunes en lecture et en écriture et l'évolution de ces compétences dans le temps* [rapport établi à la demande du Haut Conseil de l'évaluation de l'école]. DEP. Récupéré sur : <<https://www.vie-publique.fr/files/rapport/pdf/044000396.pdf>>.
- CHAPOULIE, J.-M. (2005). Entre le lycée d'élite et le lycée de masse. Paul Langevin, Gustave Monod et les réformes de l'enseignement secondaire de 1936-1939 et de 1944-1951. Dans P. Caspard, J.-N. Luc et P. Savoie (dir.), *Lycées, lycéens, lycéennes, deux siècles d'histoire* (p. 145-157). INRP. Récupéré sur : <www.persee.fr/doc/inrp_0000-0000_2005_act_28_1_9247>.
- CHERVEL, A. (2006). *Histoire de l'enseignement du français du XVII^e au XX^e siècle*. Retz.
- CHERVEL, A. (1995). *L'enseignement du français à l'école primaire. Textes officiels. Tome 2 : 1880-1939*. INRP/Economica.
- CHERVEL, A. (1993). *Histoire de l'agrégation. Contribution à l'histoire de la culture scolaire*. INRP/Kime.
- CHEVRIER, C. (2022). *Rapport au Premier ministre et à la ministre de la Transformation et de la Fonction publique*. Récupéré sur : <<https://documentation.insp.gouv.fr/insp/doc/SYRACUSE/392639/rapport-au-premier-ministre-et-a-la-ministre-de-la-transformation-et-de-la-fonction-publiques-annee->>>.
- COMITÉ NATIONAL D'ÉVALUATION. (1987). *Où va l'Université ? Rapport du Comité national d'évaluation*. Gallimard.
- DAUNAY, B. (2003). Les discours sur l'écriture d'invention et les frontières de la discipline. *Recherches*, 39, 39-69.
- DAUNAY, B. et DUFAYS, J.-L. (dir.). (2014). *Didactique du français : du côté des élèves. Comprendre les discours et les pratiques des apprenants*. De Boeck.
- DAUPHIN, L., DIEUSAERT, P., JUZDZEWSKI, L. et MICONNET, N. (2022). Les élèves du second degré à la rentrée 2022. *Note d'information de la DEPP*, 22.39. Récupéré sur : <<https://doi.org/10.48464/ni-22-39>>.
- DELAY, C. (2019) *Baccalauréat 2021. Faire du grand oral un levier d'égalité des chances* [rapport remis à Jean-Michel Blanquer]. MENJ.
- DECROIX, J., ROSSETTI, Y. et QUESQUE, F. (2022). Les neurones miroirs, hommes à tout faire des neurosciences : analyse critique des limites méthodologiques et théoriques. *L'Année psychologique*, 122(1), 85-125. Récupéré sur : <<https://doi.org/10.3917/anpsy1.221.0085>>.

- DELCAMBRE, I. (2007). Évaluation. Dans Y. Reuter (dir.), *Dictionnaire des concepts fondamentaux des didactiques* (p. 105-105). De Boeck.
- DENIZOT, N. (2005). L'écriture d'invention au lycée : écriture des genres ou nouveau genre scolaire ? *Pratiques*, 127-128, 41-44.
- DE KETELE, J.-M. (2009). L'évaluation de l'école par les standards internationaux. *Revue internationale d'éducation de Sèvres*, 52, 69-76. Récupéré sur : <<https://doi.org/10.4000/ries.743>>.
- DE PIETRO, J.-F., ROTH, M. et SANCHEZ ABCHI, V. (2016). Vers un référentiel pour l'évaluation : choix politiques et enjeux didactiques. Dans E. Falardeau, J. Dolz, J.-L. Dumortier et P. Lefrançois (dir.), *L'évaluation en classe de français, outil didactique et politique* (p. 111-132). Presses universitaires de Namur.
- DOQUET, C. et PILORGÉ, J.-L. (2020). La correction de copies au collège entre langue et discours : une catégorisation syntactico-énonciative. *Repères*, 62, 191-213. Récupéré sur : <<https://doi.org/10.4000/reperes.3274>>.
- DUPRIEZ, V. et VERHOEVEN, M. (2013). Du droit à l'éducation à l'égalité des résultats. Dans M. Frenay et X. Dumay (dir.), *Un enseignement démocratique de masse. Une réalité qui reste à inventer* (p. 19-44). Presses universitaires de Louvain.
- DUVIN-PARMENTIER, B., NOYES-ROCACHÉ, M. et GARCIA-DEBANC, C. (2021). L'évaluation des écrits dans tous ses états de vulgarisation. *Repères*, 63, 85-104. Récupéré sur : <<http://journals.openedition.org/reperes/4164>>.
- FALARDEAU, E., DOLZ, J., DUMORTIER, J.-L. et LEFRANÇOIS, P. (dir.). (2016). *L'évaluation en classe de français, outil didactique et politique*. Presses universitaires de Namur.
- GAGNON, R., BOURHIS, V. et BOURDAGES, R. (2019). Oral et évaluation : se sortir d'une dualité contradictoire ? *Pratiques*, 183-184 [en ligne]. Récupéré sur : <<https://doi.org/10.4000/pratiques.7791>>.
- GAGNON, R., MONNIER, A. et DOLZ, J. (2014). L'évaluation dans la formation à l'enseignement de la production écrite ; l'exemple suisse romand. *Pratiques*, 161-162 [en ligne]. Récupéré sur : <<https://doi.org/10.4000/pratiques.2120>>.
- GARCIA-DEBANC, C. (1999). Évaluer l'oral. *Pratiques*, 103-104, 193-212. Récupéré sur : <<https://doi.org/10.3406/prati.1999.1867>>.
- GENTAZ, É. (2022). La frénésie française des « évaluations » dans l'éducation : 4 principes pour favoriser la confiance de la communauté éducative dans leurs finalités, utilités et efficacités. *ANAE*, 180, 545-548.
- GROUPE EVA. (1996). *De l'évaluation à la réécriture*. INRP/Hachette éducation.
- GROUPE EVA. (1991). *Évaluer les écrits à l'école primaire*. INRP/Hachette éducation.
- HADJI, C. (1989). *L'évaluation, règles du jeu*. ESF.
- HOUDART-MEROT, V. et PETITJEAN, A.-M. (2018). La suppression du sujet d'invention à l'épreuve anticipée du baccalauréat. Une décision à contrecourant de l'histoire. *Le français aujourd'hui* 202(3), 105-116. Récupéré sur : <<https://www.cairn.info/revue-le-francais-aujourd-hui-2018-3-page-105.htm?contenu=resume>>.
- KINGSBURY, F. et TREMBLAY, J.-Y. (2008). *Les déterminants de l'évaluation de la langue. À l'enseignement collégial, pourquoi et jusqu'où évaluer la compétence langagière des étudiants dans les cours de la formation spécifique ?* [rapport de recherche PAREA]. Cégep de Sainte-Foy.
- KINTSCH, W. (1998). *Comprehension: A Paradigm for Cognition*. Cambridge University Press.

- LAFONTAINE, L., DUMAIS, C. et PHARAND, J. (2016). L'oral au 1^{er} cycle de l'école primaire québécoise : assises théoriques et démarche d'enseignement et d'évaluation. *Repères*, 54, 101-119. Récupéré sur : <<https://doi.org/10.4000/reperes.1096>>.
- LANDRY, F. (2006). *Structures de l'enseignement : Suisse romande et Tessin, Belgique, France, Luxembourg et Québec. Année scolaire 2006-2007*. IRDP.
- LEFRANÇOIS, P. et BRISSAUD, C. (2015). Les exigences linguistiques de quelques systèmes scolaires : une comparaison internationale. *Mesure et évaluation en éducation*, 38(3), 123-151. Récupéré sur : <<https://doi.org/10.7202/1036701ar>>.
- LEUENBERGER, M. et SARTORI, D. (2015). *Rapport final de la direction de projet et de la CESFG portant sur le sous-projet 1 Définition des compétences disciplinaires de base requises pour les études universitaires* [rapport interne]. CDIP.
- M'BAFOUMOU, A., PAC, S. et THUMERELLE, J. (2023). En fin de troisième, près de deux élèves sur trois ont une maîtrise satisfaisante des compétences numériques. *Note d'information de la DEPP*, 23.45. Récupéré sur : <<https://doi.org/10.48464/ni-23-45>>.
- MAILLARD, F. (2021). *Un diplôme sinon rien ? Les politiques de certification et leurs effets*. Dans P. Guibert (dir.), *Manuel de sciences de l'éducation et de la formation* (p. 349-356). De Boeck Supérieur. Récupéré sur : <<https://doi.org/10.3917/dbu.guibe.2021.01.0349>>.
- MAUROUX, F. et GARCIA-DEBANC, C. (2013). Analyse d'épreuves pour évaluer les compétences de scripteur des élèves à l'entrée de l'école élémentaire. *Repères*, 47, 149-170. Récupéré sur : <<https://doi.org/10.4000/reperes.535>>.
- MATHIOT, P. (2018) *Un nouveau baccalauréat pour construire le lycée des possibles* [rapport remis à Jean-Michel Blanquer, ministre de l'Éducation nationale]. MEN.
- MENOUAR, É. et CHABANNE, J. (2022). Présentation du numéro. « Évaluations et Didactiques ». *Recherches en didactiques*, 33(1), 5-9. Récupéré sur : <<https://doi.org/10.3917/rdid1.033.0005>>.
- MEURIOT, P. (1919). Le baccalauréat. Son évolution historique et statistique des origines (1808) à nos jours. *Journal de la société statistique de Paris*, 60, 8-35.
- MINISTÈRE DE L'ÉDUCATION NATIONALE ET DE LA JEUNESSE (MENJ). (2024). *Les épreuves du concours externe, du troisième concours et du second concours interne de recrutement de professeurs des écoles*. Devenir enseignant. Récupéré sur : <<https://www.devenirenseignant.gouv.fr/les-epreuves-du-concours-externe-du-troisieme-concours-et-du-second-concours-interne-de-recrutement-1009#item1>>.
- MONNIER, É. (1992). Chapitre I – La rationalisation des décisions publiques. Dans *Évaluations de l'action des pouvoirs publics* (p. 7-27). Economica.
- MONNIER-SILVA, A.-C., SCHUMACHER, T. et WEISS, L. (2020). Les examens de maturité de français langue première en Suisse sous la loupe. *Forumlecture.ch*, 3/2020 [en ligne]. Récupéré sur : <<https://www.forumlecture.ch/lfl/2020/3/706>>.
- MONS, N. (2008). Évaluation des politiques éducatives et comparaisons internationales. *Revue française de pédagogie*, 164, 5-13. Récupéré sur : <<https://doi.org/10.4000/rfp.1985>>.
- MOTTIER LOPEZ, L. (2009). L'évaluation en éducation : des tensions aux controverses. Dans L. Mottier Lopez et M. Crahay (dir.), *Évaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes* (p. 7-25). De Boeck Supérieur.
- MOTTIER LOPEZ, L. et FIGARI, G. (dir.). (2012). *Modélisations de l'évaluation en éducation. Questionnements épistémologiques*. De Boeck Supérieur.
- MOTTIER LOPEZ, L., SERRY, S. et SALES CORDEIRO, G. (2016). Évaluation certificative de productions écrites d'élèves : référence au modèle didactique du genre textuel dans l'épreuve et lors de la

- correction de faits divers. Dans E. Falardeau, J. Dolz, J.-L. Dumortier et P. Lefrançois (dir), *L'évaluation en classe de français, outil didactique et politique* (p. 27-61). Presses universitaires de Namur.
- MUELLER, S. et DARDANELLI, P. (2014). Langue, culture politique et centralisation en Suisse. *Revue internationale de politique comparée*, 21(4), 83-104. Récupéré sur : <<https://doi.org/10.3917/ripc.214.0083>>.
- NONNON, É. (2005). Entre description et prescription, l'institution de l'objet : qu'évalue-t-on quand on évalue l'oral ? *Repères*, 31, 161-188. Récupéré sur : <<https://doi.org/10.3406/reper.2005.2667>>.
- NONNON, É. (2008). La maîtrise de la langue : comment définir un « minimum commun » ? Dans D. Dubois-Marcoïn et C. Tauveron (dir.), *Français, langue et littérature, socle commun. Quelle culture pour les élèves ? Quelle professionnalité pour les enseignants ?* (p. 197-206). INRP.
- NONNON, É. (2011). Dilemmes et leçons de la construction d'un cursus de master : le poids de l'évaluation dans la définition des savoirs en formation. *Repères*, 44, 41-73. Récupéré sur : <<http://journals.openedition.org/reperes/175>>.
- PÉCAUT, F. (1911). Examen. Dans F. Buisson (dir), *Nouveau dictionnaire de pédagogie et d'instruction primaire* (édition électronique Ifé). Librairie Hachette.
- PERRIN-DOUCEY, A. (2020). Lecture littéraire et compréhension : TALC au prisme de PIRLS. Dans B. Louichon (dir.), *Un texte dans la classe. Pratiques d'enseignement de la littérature au cycle 3 en France* (p. 251-275). Peter Lang.
- PETITJEAN, A. (2003). Histoire de l'écriture d'invention au lycée. *Pratiques*, 117-118, 181-207. Récupéré sur : <<https://doi.org/10.3406/prati.2003.2002>>.
- PETITJEAN, A. (2005). Écriture d'invention au lycée et acquisition de savoir et de savoir-faire. *Pratiques*, 127-128, 75-96. Récupéré sur : <<https://doi.org/10.3406/prati.2005.2078>>.
- PETITJEAN, B. (dir.). (1984). L'évaluation [numéro thématique]. *Pratiques*, 44. Récupéré sur : <https://www.persee.fr/issue/prati_0338-2389_1984_num_44_1>.
- PLANE, S. (2022). Continuités et discontinuités dans l'enseignement de l'écriture. *Recherches*, 77, 107-129.
- PLUMELLE, B. (2002). Le système éducatif québécois. *Revue internationale d'éducation de Sèvres*, 29, 136-145. Récupéré sur : <<https://doi.org/10.4000/ries.1931>>.
- RÉMOND, M. (2005). Regards croisés sur les évaluations institutionnelles. *Repères*, 31, 113-140. Récupéré sur : <<https://doi.org/10.3406/reper.2005.2664>>.
- REUTER, Y. (2005). L'écriture d'invention : réflexions didactiques sur une réforme en cours. *Pratiques*, 127-128, 7-16. Récupéré sur : <<https://doi.org/10.3406/prati.2005.2072>>.
- SPAËTH, V. (2015). Genèse, enjeux et configuration de l'option FLES au Capes de Lettres modernes. *Le français aujourd'hui*, 189, 103-116. Récupéré sur : <<https://doi.org/10.3917/lfa.189.0103>>.
- THÉLOT, C. (2008). La principale faiblesse de l'évaluation et du contrôle de l'école en France. *Revue internationale d'éducation de Sèvres*, 48, 97-108. Récupéré sur : <<https://journals.openedition.org/ries/443>>.
- THÉLOT, C. (1994). L'évaluation du système éducatif français. *Revue française de pédagogie*, 107, 5-28. Récupéré sur : <<https://doi.org/10.3406/rfp.1994.1261>>.
- TROSSEILLE, B. et ROCHER, T. (2015). Les évaluations standardisées des élèves. Perspective historique. *Éducation et formations*, 86-87, 15-35.

- VAN BEVEREN, J., DUMORTIER, J.-L. et DISPY, M. (2013). De futurs maîtres face aux copies des lycéens. L'évaluation de la qualité de la langue dans la correction des travaux écrits. *Le français aujourd'hui*, 181(2), 95-105. Récupéré sur : <<https://doi.org/10.3917/lfa.181.0095>>.
- VANTOUROUT, M. et GOASDOUÉ, R. (2014). Approches et validité psycho-didactiques des évaluations. *Éducation et formation*, e-302, 139-156. Récupéré sur : <<http://revueeducationformation.be/index.php?revue=20&page=3>>.
- VERNEUIL, Y. (2016). La création du CAPES : révolution ou innovation contrariée (1950-1952) ? *Carrefours de l'éducation*, 41, 81-98. Récupéré sur : <<https://doi.org/10.3917/cdle.041.0081>>.
- WALINE, M. (1957). Les concours de recrutement de la fonction publique, en droit français. *International Review of Administrative Science*, 23(3), 281-292.