



HAL
open science

Qu'est-ce qu'un bon lycée ? Questions de mesure

Pauline Givord

► **To cite this version:**

Pauline Givord. Qu'est-ce qu'un bon lycée ? Questions de mesure. LIEPP Policy Brief, 2021, 54, 10.25647/liepp.pb.54 . hal-03375118

HAL Id: hal-03375118

<https://sciencespo.hal.science/hal-03375118>

Submitted on 14 Oct 2021

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.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Qu'est-ce qu'un bon lycée ? Questions de mesure

Pauline Givord*

pauline.givord@gmail.com

Au moment de la rédaction de ce *Policy brief*, chargée de mission à la direction des études et synthèses économiques à l'Insee, affiliée à l'axe Politiques éducatives du LIEPP

RÉSUMÉ

Mesurer la capacité d'un établissement scolaire à faire progresser les élèves qu'il scolarise est un exercice compliqué. Les taux de réussite au baccalauréat n'informent que peu sur, par exemple, la capacité d'un lycée à accompagner l'ensemble de leurs élèves, à leur permettre de construire un projet d'orientation et à leur donner les moyens d'y parvenir. Surtout, les résultats d'un lycée sont en très grande partie liés au niveau initial des élèves : des taux de réussite élevés peuvent traduire le degré de sélectivité du recrutement plus que la qualité du suivi des élèves. Enfin, des indicateurs sur les taux de réussite moyen d'un établissement ne renseignent qu'imparfaitement sur la réussite de l'ensemble de ses élèves. Une étude menée à partir des résultats au baccalauréat 2015 montre ainsi que, en tenant compte de la composition sociale et scolaire des établissements, certains lycées tendent à amplifier les écarts de niveau scolaire entre les élèves qu'ils scolarisent, tandis que d'autres parviennent à réduire les écarts de réussite, sans sacrifier le niveau moyen.

ABSTRACT

Measuring how a school is effective in improving students' learning is a difficult task. For example, the success rates at an exam tell us little about a school's ability to support all of its students, to assist them in identifying their educational and career aspirations and to give them the means to achieve them. A school success rate is mostly explained by the initial level of students: high success rates of a school may reflect its selectivity in enrollment more than its ability to make students improve their performance. In addition, an average success rate indicator provides only partial information about the effect on the learning improvement of every student of a school. A study based on the results of the French 2015 baccalaureate shows that, taking into account the social and academic composition of schools, some highschools tend to increase the gaps in academic level between the students they enroll, while others manage to reduce the gaps in performance between students without downgrading the average level.

Introduction

Qu'est-ce qu'un "bon" lycée ? Selon la personne à qui s'adresse cette question, la réponse a toutes les chances d'être différente. Certes, pour la plupart des parents, un bon lycée est un établissement capable d'accompagner leurs enfants jusqu'au baccalauréat, en leur assurant une scolarité sereine tout en les préparant au mieux pour l'avenir. Néanmoins, l'appréciation de la manière dont un établissement répond à ces objectifs peut varier selon les élèves. Certains adolescents s'épanouiront dans des établissements encourageant l'émulation et l'exigence académique, quand d'autres souffriront d'une ambiance trop compétitive. Au-delà des performances académiques, des parents peuvent valoriser la capacité des enseignants à développer le goût d'apprendre et la confiance en soi des élèves, la qualité du climat scolaire, ou l'aide apportée aux élèves pour construire leur orientation future et pour la rendre possible.

* L'auteure adhère à la charte de déontologie du LIEPP disponible en ligne et n'a déclaré aucun conflit d'intérêt potentiel.

Comment citer cette publication :

Pauline Givord, **Qu'est-ce qu'un bon lycée ? Questions de mesure**, LIEPP Policy Brief, n°54, 2021-10-10.

Même lorsqu'on s'accorde sur des critères de qualité des établissements, identifier l'établissement qui y répond le mieux est encore plus complexe. En général, les parents ne disposent que de peu d'éléments pour juger d'un établissement. Les expériences passées de connaissances ou de la fratrie, le taux de réussite au baccalauréat sont des informations certes utiles, mais qui ne renseignent qu'indirectement sur la manière dont un élève particulier s'adaptera finalement à un lycée. Disposer de mesures des effets établissements est pourtant indispensable pour permettre de juger de l'action effective des équipes pédagogiques, au-delà des effets de réputation. Mais définir des indicateurs pertinents n'est pas immédiat. La réussite affichée par un lycée est avant tout le reflet de son recrutement, et ne pas tenir compte de ces effets de sélection peut conduire à donner des images incorrectes de la qualité des établissements et donc des informations peu pertinentes pour les familles. Par exemple, un lycée qui sélectionne ses élèves en fonction de leur dossier scolaire est presque certain d'afficher un taux de réussite au baccalauréat très élevé. Cela ne signifie pas que tout élève qui serait scolarisé dans un tel établissement, quel que soit son niveau de départ, sera assuré d'obtenir d'aussi bons résultats. Cela appelle à des indicateurs qui permettent non seulement de tenir compte des effets de composition des lycées, mais qui donnent également une indication de ce qui se passe "au-delà de la moyenne".

Modéliser les effets établissements et non la qualité de leur recrutement

Dans les dernières décennies, de nombreux travaux de recherche se sont penchés sur la mesure les effets des établissements sur la réussite des élèves. Le problème principal est celui des effets de sélection, lié au fait que les lycées scolarisent en général des élèves dont les caractéristiques personnelles diffèrent, et que ces caractéristiques (niveau scolaire initial, implication des parents, etc.) sont plus ou moins propices à la réussite. Pour mesurer l'efficacité relative des lycées entre eux, il faudrait, en théorie, affecter aléatoirement des élèves entre les lycées, ce qui est évidemment impossible.

Plusieurs types de modèles ont été développés pour réduire les biais liés à ces effets de compositions, mais deux principaux prédominent : les modèles de Student Growth Percentile, et les modèles de valeurs ajoutées. Les modèles de Student Growth Percentile ont été développés initialement par l'Etat américain du Colorado pour évaluer l'efficacité des établissements voire des enseignants, et leur usage s'est diffusé depuis. Leur principe consiste à "classer" les élèves selon leurs résultats à des tests, le rang dans ce classement étant représenté par le percentile dans la distribution des notes. Par exemple, si un élève fait mieux à un test de fin d'année que 80% des élèves qui avaient en début d'année un niveau proche du sien, on attribuera au lycée un effet positif de 80 pour cet élève. L'efficacité de l'établissement correspond à la moyenne de ces effets mesurés pour l'ensemble des

élèves qui y sont scolarisés. Néanmoins, tenir compte uniquement du niveau initial ne réduit pas totalement les problèmes de sélection mentionnés plus haut. D'autres facteurs, comme l'origine sociale par exemple, peuvent influencer sur la réussite scolaire. Ces indicateurs ont tendance à pénaliser les établissements qui seraient face à des élèves issus de milieu social défavorisé ou avec des besoins particuliers (Walsh et Isenberg, 2015).

L'autre grande famille de modèles pour mesurer les effets établissements repose sur l'estimation de la "valeur ajoutée" des lycées. Ils consistent à modéliser les indicateurs de réussite scolaire en fonction d'un certain nombre de caractéristiques observables, comme le niveau initial ou encore le milieu d'origine, ainsi qu'un effet propre à l'établissement. Cet effet propre correspond à l'écart entre ce qui est attendu, compte tenu des caractéristiques observables des établissements, et ce qui est effectivement observé en moyenne pour les élèves de l'établissement. C'est en particulier ce type de modèles qui est utilisé par la Direction de l'évaluation, de la prospective et de la performance (Depp) du ministère de l'Education Nationale, pour mesurer les Indicateurs de Valeur Ajoutée (les IVAL) des lycées, publiés chaque année (Evain et Evrard, 2017).

Aller au-delà de la moyenne : l'excellence pour tous ?

Cependant, la plupart des analyses disponibles pour mesurer les effets lycées se concentrent sur des effets moyens, ce qui peut masquer des pratiques très différentes entre lycées, et donc des résultats individuels très variables au sein d'un même établissement. Prenons l'exemple fictif de deux lycées, accueillant chacun trois élèves de niveau identique à l'origine. Dans le premier, à l'issue de leur scolarité, tous les élèves obtiennent la note de 12/20. Dans le deuxième lycée à l'inverse, l'un des élèves obtient 18/20 et les deux autres 9/20 : si en moyenne, les résultats seront identiques au premier lycée, ici les très bons résultats d'un élève s'accompagnent du décrochage des deux autres. Selon l'un ou l'autre cas, l'indicateur moyen de l'effet lycée donnera une information dont la précision sera très différente. Si pour le premier lycée, le résultat "espéré" est relativement sûr, mais s'établit à un niveau plutôt médiocre, l'effet attendu sur un élève est beaucoup plus "ouvert", avec une probabilité de bonnes ou de mauvaises surprises. Il importe donc d'étudier les effets des établissements "au-delà" de la moyenne.

Les données de la Depp permettent de construire des indicateurs complémentaires aux mesures de la performance des lycées, à partir des résultats obtenus par les lycéens des séries générale et technologique de la session 2015 du baccalauréat (voir encadré). La valeur ajoutée des lycées est estimée non seulement au niveau de la médiane, mais également aux niveaux respectivement des plus faibles et des meilleurs élèves. La méthode consiste à estimer l'effet propre de chaque lycée sur le niveau atteint par au moins 80% des élèves, ainsi que sur le

Encadré : Estimation des effets fixes des lycées par des régressions quantiles

Cette étude (Givord et Suarez, 2020) s'appuie sur la base de données exhaustive des résultats à l'examen national du baccalauréat de la session 2015, complétés par le système d'information des « Fichiers anonymisés pour les études et la recherche » (Faere) produit par la DEPP, qui contient des informations individuelles sur les élèves et notamment les résultats individuels à l'examen national du brevet (DNB), qui constitue un indicateur du niveau scolaire de l'élève au moment de l'entrée au lycée. Les estimations sont menées pour les lycées généraux et technologiques, en utilisant des modèles distincts pour les deux filières.

On estime les effets spécifiques du lycée où est scolarisé, une fois tenu compte des autres caractéristiques individuelles des élèves (résultats moyens obtenus à l'examen du brevet, avec une forme quadratique, origine sociale, fait d'avoir redoublé ou pas, statut vis-à-vis de l'immigration, genre), sur plusieurs niveaux de la distribution de la note moyenne obtenue au baccalauréat : premier quintile (c'est-à-dire le niveau tel que seulement un cinquième des élèves ont un niveau inférieur), médiane (le niveau tel que la moitié des élèves ont un niveau inférieur) et dernier quintile (le niveau tel que un cinquième des élèves ont un niveau supérieur). Si la médiane est en général très proche du niveau moyen, les premier et dernier quintiles renseignent sur les niveaux respectivement des élèves plus en difficultés et des plus brillants. L'estimation se fait à partir d'une méthode de régression quantile, qui suppose une relation linéaire entre le quantile d'ordre τ (compris entre 0 et 1, ici τ prend les trois valeurs 0,2, 0,5 et 0,8) de la distribution de la variable d'intérêt Y conditionnelle aux caractéristiques de l'élève X scolarisé dans le lycée j , soit $q_{\tau}(X) = X\beta_{\tau} + \alpha_{\tau}$.

Le coefficient α_{τ} correspond à la manière dont le quantile d'ordre τ des notes moyennes obtenues par les élèves de ce lycée s'écarte de ce qui est observé sur l'ensemble des lycées, en tenant compte des caractéristiques observables individuelles des élèves. Par exemple, $\alpha_{0,8j}$ correspond à l'écart estimé entre le niveau tel que seuls 20% des élèves du lycée j ont obtenu une note supérieure au baccalauréat, et le niveau qui est attendu pour seulement 20% des élèves de niveau initial et de caractéristiques individuelles observables identiques.

En pratique, comparer pour chaque lycée l'écart entre les deux vingtiles extrêmes $\alpha_{0,8}^j - \alpha_{0,2}^j$, ce qui renseigne sur la différence de dispersion dans ce lycée, par rapport à ce qui est attendu sur l'ensemble des lycées, une fois pris en compte la composition. Un test statistique permet de déterminer les lycées pour lesquels cette différence est significativement positive, qui correspond aux lycées qui ont tendance à accroître la dispersion des notes, par rapport à ce qui est observé sur l'ensemble des lycées, et ceux pour lesquels cette différence est significativement négative, qui correspond aux lycées qui à l'inverse ont tendance à réduire les écarts.

niveau atteint par moins de 20% des élèves, en tenant compte des caractéristiques des élèves qui sont observées dans les données, comme les résultats obtenus à l'examen du brevet, l'origine sociale, le fait d'avoir redoublé ou pas, etc.

Il est ainsi notamment possible d'identifier des lycées qui ont tendance à réduire les écarts de réussite entre les élèves les plus faibles et les meilleurs dans le lycée, par rapport à ce qui serait attendu, ou à l'inverse ceux qui ont tendance à augmenter ces écarts de réussite entre élèves du lycée. Si, pour la majorité des lycées, il n'est pas possible de mettre en évidence statistiquement des effets hétérogènes – les écarts observés sont du même ordre que les écarts attendus, environ un sixième d'entre eux ont tendance soit à amplifier, soit à résorber les écarts entre les résultats de leurs élèves.

Ces indicateurs peuvent être comparés à la valeur ajoutée des mêmes lycées, estimée au niveau de la médiane (qui est en général proche de la moyenne). La réduction de la dispersion des

résultats des étudiants n'est en effet pas un objectif en soi. Elle n'est pas souhaitable si elle doit aboutir à un nivellement par le bas, c'est-à-dire si les résultats moins inégaux entre les élèves se font au détriment du niveau moyen. Mais cette crainte ne semble pas fondée. On constate en effet que la performance moyenne est en général supérieure dans les lycées qui réduisent les inégalités, comme illustré dans le graphique où on observe une relation négative entre performance au niveau de la médiane et dispersion des résultats que ce soit dans les séries générales ou technologiques. Les lycées dont la dispersion en terme de valeur ajoutée est la plus faible (les points gris, situés vers le bas de l'axe vertical) sont relativement plus nombreux parmi ceux qui réussissent le mieux en moyenne (vers la droite de l'axe horizontal) par rapport aux lycées qui tendent à augmenter la dispersion des résultats (les lycées situés vers le haut de l'axe vertical).

Cela signifie que certains lycées parviennent à faire réussir leurs élèves sans sacrifier les plus faibles.

Effets lycées sur la dispersion et la médiane (voies générale et technologique)



Source : MENJ-DEPP, fichiers Faere

Parmi les lycées qui réduisent les écarts entre élèves dans la voie générale, 46% ont également un effet significativement positif sur les résultats moyens. A l'inverse, toujours dans la voie générale, 31% des lycées qui augmentent les inégalités entre les élèves sont classés comme peu performants, au sens où la valeur ajoutée au niveau de la moyenne est significativement négative. Cependant, cette relation entre efficacité et égalité n'est pas univoque. Certains lycées "égalitaires" qui parviennent à réduire les écarts de performance entre leurs élèves, sont parfois moins performants : c'est le cas de 15% d'entre eux, tandis qu'une proportion similaire de lycées "inégalitaires" ont aussi des résultats moyens supérieurs à ce qui est attendu.

Que retenir ?

Un premier message est que certains établissements semblent à la fois capables de faire progresser l'ensemble de leurs élèves, sans que cela se traduise par un nivellement vers le bas. Mais plus généralement, cette étude suggère qu'un indicateur unique ne peut permettre de donner une image correcte de l'effet d'un établissement. Cela ne signifie pas qu'il faille renoncer à utiliser des indicateurs quantitatifs, mais qu'il est indispensable de croiser les dimensions pour capter cette complexité.

C'est d'ailleurs ce que fait en France la Depp. Sont publiés chaque année des indicateurs sur le parcours des élèves au sein des établissements, sur la réussite au baccalauréat ou encore sur le taux de mentions. Il est également possible d'envisager d'autres dimensions, à l'image de ce qui est utilisé dans certains pays. Par exemple, en Italie, le projet Eduscopio fournit, à partir de l'appariement avec des sources administratives, des évaluations des établissements scolaires du second degré à partir d'indicateurs portant sur la réussite dans le supérieur ou sur le marché du travail (Aina et al., 2021). De fait, les épreuves terminales du baccalauréat ont de moins en moins d'incidence sur l'accès à l'enseignement supérieur et ne présagent pas de la réussite ultérieure et de l'accès à l'emploi.

Par ailleurs, si les évaluations disponibles tendent à se focaliser sur ce qu'on sait le mieux mesurer, la réussite des élèves à des tests scolaires, il serait pertinent de s'intéresser à des compétences plus transversales mais peut être au moins aussi importantes, comme la capacité des enseignants à développer chez leurs élèves la confiance en soi, l'envie d'apprendre ou l'esprit critique, dimensions qui ne se recoupent que partiellement avec les compétences cognitives. Par exemple, des études américaines récentes montrent que les capacités des enseignants à faire réussir leurs élèves à des tests standardisés sont très peu corrélées à leur capacités à développer l'effort et la motivation des élèves (Kraft, 2019) ou à agir sur l'absentéisme ou le redoublement, des dimensions pourtant prédictives de la réussite scolaire sur le long terme (Jackson, 2018). Cela suggère que des mesures exclusivement focalisées sur la performance scolaire ne fournissent qu'une image incomplète de la qualité réelle des établissements, et

plaide pour compléter les évaluations des établissements par des mesures plus qualitatives, par exemple sur des activités ou des méthodes pédagogiques permettant de développer ces compétences transversales.

Une autre question essentielle est celle de l'appropriation de ces indicateurs par les familles: de fait, une étude sur les choix des parents à New York (Abdulkadiroğlu et al., 2020) montre que les parents valorisent essentiellement les lycées qui scolarisent les élèves avec les meilleurs niveaux scolaires, quelle que soit l'efficacité réelle de ces établissements à faire progresser leurs élèves. Cela, alors même que pour certains élèves, être scolarisés dans ces établissements au recrutement élitiste peut avoir un effet négatif sur la réussite future (Abdulkadiroğlu et al., 2014). En retour, cela signifie que les établissements dont la réputation est déjà établie ont peu d'incitations à miser sur la qualité, et peuvent se contenter de sélectionner les meilleurs élèves, entretenant une ségrégation scolaire de fait, entre les établissements les plus attractifs et d'autres scolarisant des élèves en difficulté. Assurer une meilleure information des familles sur les effets réels des lycées peut faire partie des outils permettant de réduire ces fractures scolaires, dont les effets peuvent être délétères sur l'équité et aussi l'efficacité scolaire (Givord, 2019).

Références

- ABDULKADIROGLU, A. et al., P., 2014, The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica*, 82: 137-196.
- ABDULKADIROGLU, A., et al., 2020, Do Parents Value School Effectiveness?, *American Economic Review*, 110 (5): 1502-39.
- AINA, C. et al., 2021, Ranking high schools using university student performance in Italy, *Economica Politica* 38, 293-321.
- EVAIN, F., EVRARD, I., 2017, Une meilleure mesure de la performance des lycées: Refonte de la méthodologie des IVAL (session 2015). *Education et Formations*, Vol. 94, pp. 91-116.
- GIVORD, P., 2019, Does greater social diversity in schools have an impact on equity in learning outcomes?, *PISA in Focus*, n°97, Éditions OCDE, Paris.
- GIVORD, P., SUAREZ CASTILLO, M., 2020, Excellence for all ? Heterogeneity in high schools' value-added, *LIEPP Working Paper* n°111.
- GUARINO, C. et al., 2015, A Comparison of Student Growth Percentile and Value-Added Models of Teacher Performance, *Statistics and Public Policy*, Vol. 2/1, pp. 1-11.
- JACKSON, C., 2018, What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes, *Journal of Political Economy*, Vol. 126/5, pp. 2072-2107.
- KOEDIGER, C., MIHALY, K., ROCKOFF J., 2015, Value-added modeling: A review, *Economics of Education Review*, Vol. 47, pp. 180-195.
- KRAFT, M., 2019, Teacher Effects on Complex Cognitive Skills and Social-Emotional Competencies, *Journal of Human Resources*, Vol. 54/1, pp. 1-36.
- KURTZ, M., 2018, Value-Added and Student Growth Percentile Models: What Drives Differences in Estimated Classroom Effects?, *Statistics and Public Policy*, Vol. 5/1, pp. 1-8.
- ROTHSTEIN, J., 2010, Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement, *Quarterly Journal of Economics*, Vol. 125/1, pp. 175-214.
- SASS, T. et al., 2014, Value-added models and the measurement of teacher productivity, *Economics of Education Review*, Vol. 38, pp. 9-23
- WALL, D., 2000, The impact of high-stakes testing on teaching and learning: can this be predicted or controlled?, *System*, Vol. 28/4, pp. 499-509.
- WALSH, E., ISENBERG, E., 2015, How Does Value Added Compare to Student Growth Percentiles?, *Statistics and Public Policy*, Vol. 2/1, pp. 1-13.



Le LIEPP (Laboratoire interdisciplinaire d'évaluation des politiques publiques) est un laboratoire d'excellence (Labex) distingué par le Jury scientifique international désigné par l'Agence nationale de la recherche (ANR). Il est financé dans le cadre des investissements d'avenir (ANR-11-LABX-0091, ANR-11-IDEX-0005-02) et de l'IdEx Université de Paris (ANR-18-IDEX-0001).

Directrice de publication : Anne Revillard

Édition et maquette : Sofia Cerda Aparicio, Andreana Khristova

Sciences Po - LIEPP
27 rue Saint Guillaume
75007 Paris - France
© LIEPP 2021

www.sciencespo.fr/liepp

@LIEPP_ScPo