HAL open science

Comment évaluer le succès d'un dispositif destiné à améliorer les résultats scolaires ou pourquoi il est impératif de trouver d'autres indicateurs que les notes ?

Muriel Epstein

▶ **To cite this version:**

Muriel Epstein. Comment évaluer le succès d'un dispositif destiné à améliorer les résultats scolaires ou pourquoi il est impératif de trouver d'autres indicateurs que les notes ?. Congrès Intaref 2013 - Actualité de la recherche en education et en Formation - , LIRDEF, Aug 2013, Montpellier, France. hal-01364602

HAL Id: hal-01364602

https://hal.science/hal-01364602v1

Submitted on 12 Sep 2016

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 - NonCommercial - ShareAlike 4.0 International License

Ads by PlusHD.5 Ad Options

251 - Comment évaluer le succès d'un dispositif destiné à améliorer les résultats scolaires ou pourquoi il est impératif de trouver d'autres indicateurs que les notes ?

Muriel Epstein

Transpi, France

Mots clés : évaluation, dispositifs, notes, indicateurs, gouvernance, méthodologie

Résumé : Comme la loi l'impose, les dispositifs pour aider les élèves en difficultés sont systématiquement évalués afin de savoir comment l'argent public a été dépensé mais également pour connaître leur efficacité, l'opportunité d'élargir ou de poursuivre l'expérience, etc. Et, sauf pour quelques rares actions, la question récurrente des pouvoirs publics reste l'indicateur ultime à savoir « les notes ont-elles progressé ? » ou éventuellement « a-t-on amélioré le taux de réussite à l'examen final ? ».

Cette dernière question est particulièrement importante depuis les accords de Lisbonne de 2005 engageant les états européens à 80% de réussite au baccalauréat et 50% à bac +3.

Or de nombreuses évaluations réalisées par nos soins (l'une sur un dispositif public national, l'autre sur une expérimentation privée locale) ou par d'autres chercheurs (Brito, 2012 ; Cousin, 1998) montrent que les résultats d'une expérimentation sont rarement réductibles à la note, voire pire qu'en général les évaluations des notes oublient des variables (Bressoux, P., Lima, L., 2011) et ne tiennent pas compte de la dimension temporelle de l'apprentissage (Vandelle, 2012) à savoir que les résultats scolaires n'arrivent, qu'en dernier recours. En effet, le chemin de progression (qui dépend du genre de l'adolescent) passe par le renforcement du sentiment de sécurité (être protégé, faire confiance) et d'identité (savoir ce dont on est capable ou pas) puis le sentiment d'appartenance au groupe, puis le sentiment de détermination (se fixer des buts) et enfin le sentiment de compétence (atteindre ses buts). Ce n'est que lorsque le sentiment de compétence est restauré que l'amélioration des résultats scolaires devient visible, au minimum après un travail de plusieurs mois voire souvent de plusieurs années.

Après avoir détaillé les différentes expérimentations, et précisé les statuts des évaluations, en particulier celles où je suis intervenue (certaines ont été réalisées comme chercheuse, d'autres comme enseignante et d'autres enfin comme consultante) et certains résultats mis en évidence par les évaluations, dont l'importance de la motivation et de l'estime de soi, la communication propose de réfléchir aux difficultés méthodologiques et aux prismes qu'engendrent la primauté du « résultat scolaire » comme indicateur de succès d'un dispositif de lutte contre l'échec scolaire. Nous conclurons sur d'autres méthodes possibles d'évaluation.

1. Introduction

Ads by PlusHD.5 Ad Options

251 - Comment évaluer le succès d'un dispositif destiné à améliorer les résultats scolaires ou pourquoi il est impératif de trouver d'autres indicateurs que les notes ?

Muriel Epstein

Transpi, France

Mots clés : évaluation, dispositifs, notes, indicateurs, gouvernance, méthodologie

Résumé : Comme la loi l'impose, les dispositifs pour aider les élèves en difficultés sont systématiquement évalués afin de savoir comment l'argent public a été dépensé mais également pour connaître leur efficacité, l'opportunité d'élargir ou de poursuivre l'expérience, etc. Et, sauf pour quelques rares actions, la question récurrente des pouvoirs publics reste l'indicateur ultime à savoir « les notes ont-elles progressé ? » ou éventuellement « a-t-on amélioré le taux de réussite à l'examen final ? ».

Cette dernière question est particulièrement importante depuis les accords de Lisbonne de 2005 engageant les états européens à 80% de réussite au baccalauréat et 50% à bac +3.

Or de nombreuses évaluations réalisées par nos soins (l'une sur un dispositif public national, l'autre sur une expérimentation privée locale) ou par d'autres chercheurs (Brito, 2012 ; Cousin, 1998) montrent que les résultats d'une expérimentation sont rarement réductibles à la note, voire pire qu'en général les évaluations des notes oublient des variables (Bressoux, P., Lima, L., 2011) et ne tiennent pas compte de la dimension temporelle de l'apprentissage (Vandelle, 2012) à savoir que les résultats scolaires n'arrivent, qu'en dernier recours. En effet, le chemin de progression (qui dépend du genre de l'adolescent) passe par le renforcement du sentiment de sécurité (être protégé, faire confiance) et d'identité (savoir ce dont on est capable ou pas) puis le sentiment d'appartenance au groupe, puis le sentiment de détermination (se fixer des buts) et enfin le sentiment de compétence (atteindre ses buts). Ce n'est que lorsque le sentiment de compétence est restauré que l'amélioration des résultats scolaires devient visible, au minimum après un travail de plusieurs mois voire souvent de plusieurs années.

Après avoir détaillé les différentes expérimentations, et précisé les statuts des évaluations, en particulier celles où je suis intervenue (certaines ont été réalisées comme chercheuse, d'autres comme enseignante et d'autres enfin comme consultante) et certains résultats mis en évidence par les évaluations, dont l'importance de la motivation et de l'estime de soi, la communication propose de réfléchir aux difficultés méthodologiques et aux prismes qu'engendrent la primauté du « résultat scolaire » comme indicateur de

succès d’un dispositif de lutte contre l’échec scolaire. Nous conclurons sur d’autres méthodes possibles d’évaluation.

1. Introduction

expérimentation visant à améliorer les résultats scolaires des élèves d'une classe. 2012

Lorsque l'on teste un médicament, on veut savoir s'il fonctionne. En conséquence, on mesure ses effets directs et secondaires. En général, face à une maladie donnée, sont identifiés des indicateurs. Par exemple, la fièvre.

Quoi de plus logique en effet, lorsqu'un médicament vise à diminuer la température, de vérifier qu'il le fait ? Mais l'indicateur peut varier selon la maladie. Ainsi, on peut utiliser le taux de lymphocyte comme un indicateur de contamination par le VIH. Ce ne sera jamais qu'un marqueur parmi d'autres mais, à un moment donné, il est jugé suffisamment fiable pour être utilisé comme tel et fera partie de la batterie de tests visant à décider qu'un traitement est, ou non, efficace.

En sport, le principe est le même : la plupart des référentiels évaluent l'efficacité d'un sportif (et d'un entraineur sportif) par les performances de l'athlète entrainé. S'il est possible d'attendre « une saison » pour observer l'efficacité d'une stratégie, on peut mesurer en nombre de buts, en nombre de secondes ou millisecondes gagnées l'effet de la méthode mise en œuvre.

Dans différents ouvrages (par exemple Felouzis & Hanard 2011), les chercheurs et consultants évaluateurs expliquent que l'un des écueils consisterait à croire que toute évaluation est le reflet fidèle de la réalité qu'elle est censée mesurer. Un autre écueil est l'attitude de rejet de toute évaluation sous prétexte que les résultats de l'éducation échappent en grande partie au mesurable

Comme le rappelle Michel Lecointe[1], le sens le plus étroit de l'évaluation économique d'un dispositif est « quelque part entre 'qu'avez-vous fait de ma subvention ?' et l'interrogation sur le retour d'investissement. C'est, en général, la première demande des institutions territoriales : **l'objectif étant de pouvoir afficher qu'il a bien été fait quelque chose de visible et de palpable avec l'argent du contribuable** »

Demande bien légitime donc.

Nous garderons, pour cet article, une définition assez large de l'évaluation de Xavier Pons à savoir « un processus collectif et réflexif de construction d'un retour d'information sur le fonctionnement du système éducatif permettant d'apprécier, en référence à des valeurs politiques multiples (pertinence, cohérence, légalité, efficacité, efficience ou encore équité), son organisation, ses résultats et les effets de différentes mesures mises en œuvre (politiques, programmes d'action, dispositifs, projets, innovations, expérimentations »[2] .

L'école vise, en France, à éduquer (à la démocratie, au « vivre ensemble », etc.) et à transmettre des connaissances et compétences (le « socle commun[3] »). **Mesurer l'efficacité de l'école revient donc généralement à évaluer si ces connaissances et compétences sont acquises.**

Or, en première intuition, **c'est bien la note qui indique si, oui ou non, des compétences sont acquises**. C'est même la note qui définit ce qu'est l'acquisition des compétences (Antibi 2003). La compétence « savoir lire » est considérée comme acquise si l'élève a « eu la moyenne » en lecture.

En effet, le contrôle ou l'interrogation est censé indiquer tant à l'institution qu'aux parents ou aux élèves eux-mêmes s'ils ont, ou non, compris ce qui était attendu d'eux, à savoir, en général, le programme.

Un collégien qui aura obtenu plus de 16 à un contrôle est généralement perçu comme « ayant bien compris » tandis qu'un élève qui a 5 de moyenne est considéré comme « ne suivant pas ».

Si « avoir 10 de moyenne » marque théoriquement le fait de valider (un apprentissage, un passage en classe supérieure, un examen, etc…), des enseignants rappellent régulièrement[4] à quel point cette « moyenne » ne signifie rien quant aux compétences acquises et à la possibilité de poursuivre des études…. Tout en se basant sur ce même indicateur pour expliquer à l'élève ou à ses parents, les marges de progressions.

expérimentation visant à améliorer les résultats scolaires des élèves d'une classe. 2012

Lorsque l'on teste un médicament, on veut savoir s'il fonctionne. En conséquence, on mesure ses effets directs et secondaires. En général, face à une maladie donnée, sont identifiés des indicateurs. Par exemple, la fièvre.

Quoi de plus logique en effet, lorsqu'un médicament vise à diminuer la température, de vérifier qu'il le fait ? Mais l'indicateur peut varier selon la maladie. Ainsi, on peut utiliser le taux de lymphocyte comme un indicateur de contamination par le VIH. Ce ne sera jamais qu'un marqueur parmi d'autres mais, à un moment donné, il est jugé suffisamment fiable pour être utilisé comme tel et fera partie de la batterie de tests visant à décider qu'un traitement est, ou non, efficace.

En sport, le principe est le même : la plupart des référentiels évaluent l'efficacité d'un sportif (et d'un entraineur sportif) par les performances de l'athlète entrainé. S'il est possible d'attendre « une saison » pour observer l'efficacité d'une stratégie, on peut mesurer en nombre de buts, en nombre de secondes ou millisecondes gagnées l'effet de la méthode mise en œuvre.

Dans différents ouvrages (par exemple Felouzis & Hanard 2011), les chercheurs et consultants évaluateurs expliquent que l'un des écueils consisterait à croire que toute évaluation est le reflet fidèle de la réalité qu'elle est censée mesurer. Un autre écueil est l'attitude de rejet de toute évaluation sous prétexte que les résultats de l'éducation échappent en grande partie au mesurable

Comme le rappelle Michel Lecointe[1], le sens le plus étroit de l'évaluation économique d'un dispositif est « quelque part entre 'qu'avez-vous fait de ma subvention ?' et l'interrogation sur le retour d'investissement. C'est, en général, la première demande des institutions territoriales : l'objectif étant de pouvoir afficher qu'il a bien été fait quelque chose de visible et de palpable avec l'argent du contribuable »

Demande bien légitime donc.

Nous garderons, pour cet article, une définition assez large de l'évaluation de Xavier Pons à savoir « un processus collectif et réflexif de construction d'un retour d'information sur le fonctionnement du système éducatif permettant d'apprécier, en référence à des valeurs politiques multiples (pertinence, cohérence, légalité, efficacité, efficience ou encore équité), son organisation, ses résultats et les effets de différentes mesures mises en œuvre (politiques, programmes d'action, dispositifs, projets, innovations, expérimentations »[2] .

L'école vise, en France, à éduquer (à la démocratie, au « vivre ensemble », etc.) et à transmettre des connaissances et compétences (le « socle commun[3] »). Mesurer l'efficacité de l'école revient donc généralement à évaluer si ces connaissances et compétences sont acquises.

Or, en première intuition, c'est bien la note qui indique si, oui ou non, des compétences sont acquises. C'est même la note qui définit ce qu'est l'acquisition des compétences (Antibi 2003). La compétence « savoir lire » est considérée comme acquise si l'élève a « eu la moyenne » en lecture.

En effet, le contrôle ou l'interrogation est censé indiquer tant à l'institution qu'aux parents ou aux élèves

eux-mêmes s’ils ont, ou non, compris ce qui était attendu d’eux, à savoir, en général, le programme.

Un collégien qui aura obtenu plus de 16 à un contrôle est généralement perçu comme « ayant bien compris » tandis qu’un élève qui a 5 de moyenne est considéré comme « ne suivant pas ».

Si « avoir 10 de moyenne » marque théoriquement le fait de valider (un apprentissage, un passage en classe supérieure, un examen, etc...), des enseignants rappellent régulièrement[4] à quel point cette « moyenne » ne signifie rien quant aux compétences acquises et à la possibilité de poursuivre des études.... Tout en se basant sur ce même indicateur pour expliquer à l’élève ou à ses parents, les marges de progressions.

formation enseignante, etc.).

Et pourtant, même pour évaluer l'impact d'un dispositif destiné à améliorer les résultats scolaires, peut-être même surtout pour évaluer un tel impact, **il est indispensable de trouver d'autres résultats que les notes**. C'est ce que va s'employer à démontrer cet article.

Il s'agit d'une méta recherche qui s'appuiera sur trois types de matériaux :

- Pour l'essentiel, des recherches en évaluations de politiques publiques réalisées par des sociologues et des consultants dont des évaluations de dispositifs d'amélioration des résultats scolaires des élèves auxquelles j'ai participé
- Des analyses du mode de progression et de notations des élèves, dont les recherches d'Hélène Vandelle que j'ai appuyé sur le traitement quantitatif des données
- des expériences réalisées en tant qu'enseignante dans l'objectif de construire un enseignement plus efficace notamment dans le cadre de deux expériences de lutte contre le décrochage scolaire que sont Booster (qui vise à faire découvrir aux jeunes de plus de 16 ans leur passion) et Transapi (qui vise à apprendre hors les murs de l'école, sans note et sans évaluation externe des élèves)

Dans un premier temps, nous rappellerons en quoi la note, dans une classe, à un moment donné, évalue rarement un niveau scolaire. Puis nous nous intéresserons aux controverses à propos des évaluations dites « externes » (bac, PISA) généralement plébiscitées car elles permettraient l'objectivation recherchée.

Une seconde partie propose de détailler en quoi la note ne peut pas être le premier résultat d'un dispositif visant à améliorer les performances scolaires des élèves avant de démontrer, à travers deux exemples, en quoi cette approche peut s'avérer contre-productive.

Enfin, une longue partie conclusive vise à réaffirmer le besoin d'évaluer les dispositifs éducatifs, en particulier innovants, tout en réfléchissant aux mesures ou aux précautions qui peuvent être utiles ainsi qu'aux moyens possibles de mesurer et d'évaluer le succès d'un dispositif autrement que par les notes.

Cet article est la première étape, indispensable, d'une recherche sur la création d'évaluations standardisées du rapport au savoir des élèves.

2. La note comme évaluation d'un niveau ?

2.0. A quoi sert la note ?

La note a plusieurs fonctions :

- **Pronostique** : l'objectif de cette forme d'évaluation est de prévoir la réussite d'un apprenant dans une formation et, par exemple, de constituer des groupes de niveau (en langue notamment). C'est aussi à partir d'évaluations pronostiques que les enseignants orientent les élèves
- **Diagnostique** : savoir où en sont les élèves en début d'un cours, les points forts et faibles. L'évaluation diagnostique est parfois réalisée en début d'année.
- **Formative** : une note est dite formative si elle permet à un élève de se situer, et de l'aider à progresser
- **Sommative** ou Bilan : dresser un bilan des connaissances et des compétences

Comme le note l'académie de Rouen sur son site, théoriquement, seule la dernière est à visée institutionnelle.

L'évaluation à l'école oscille entre le formatif, les enseignants souhaitant former les élèves et le bilan. (Antibi 2003, Baillat, Ketele, Paquay, Thelot 2008).

formation enseignante, etc.).

Et pourtant, même pour évaluer l’impact d’un dispositif destiné à améliorer les résultats scolaires, peut-être même surtout pour évaluer un tel impact, il est indispensable de trouver d’autres résultats que les notes. C’est ce que va s’employer à démontrer cet article.

Il s’agit d’une méta recherche qui s’appuiera sur trois types de matériaux :

Pour l’essentiel, des recherches en évaluations de politiques publiques réalisées par des sociologues et des consultants dont des évaluations de dispositifs d’amélioration des résultats scolaires des élèves auxquelles j’ai participé Des analyses du mode de progression et de notations des élèves, dont les recherches d’Hélène Vandelle que j’ai appuyé sur le traitement quantitatif des données des expériences réalisées en tant qu’enseignante dans l’objectif de construire un enseignement plus efficace notamment dans le cadre de deux expériences de lutte contre le décrochage scolaire que sont Booster (qui vise à faire découvrir aux jeunes de plus de 16 ans leur passion) et Transapi (qui vise à apprendre hors les murs de l’école, sans note et sans évaluation externe des élèves)

Dans un premier temps, nous rappellerons en quoi la note, dans une classe, à un moment donné, évalue rarement un niveau scolaire. Puis nous nous intéresserons aux controverses à propos des évaluations dites « externes » (bac, PISA) généralement plébiscitées car elles permettraient l’objectivation recherchée.

Une seconde partie propose de détailler en quoi la note ne peut pas être le premier résultat d’un dispositif visant à améliorer les performances scolaires des élèves avant de démontrer, à travers deux exemples, en quoi cette approche peut s’avérer contre-productive.

Enfin, une longue partie conclusive vise à réaffirmer le besoin d’évaluer les dispositifs éducatifs, en particulier innovants, tout en réfléchissant aux mesures ou aux précautions qui peuvent être utiles ainsi qu’aux moyens possibles de mesurer et d’évaluer le succès d’un dispositif autrement que par les notes.

Cet article est la première étape, indispensable, d’une recherche sur la création d’évaluations standardisées du rapport au savoir des élèves.

2. La note comme évaluation d’un niveau ?

2.0. A quoi sert la note ?

La note a plusieurs fonctions :

Pronostique : l’objectif de cette forme d’évaluation est de prévoir la réussite d’un apprenant dans une formation et, par exemple, de constituer des groupes de niveau (en langue notamment). C’est aussi à partir d’évaluations pronostiques que les enseignants orientent les élèves Diagnostique : savoir où en sont les élèves en début d’un cours, les points forts et faibles. L’évaluation diagnostique est parfois réalisée en début d’année. Formative : une note est dite formative si elle permet à un élève de se situer, et de l’aider à progresser Sommative ou Bilan : dresser un bilan des connaissances et des compétences

Comme le note l’académie de Rouen sur son site, théoriquement, seule la dernière est à visée

institutionnelle.

L’évaluation à l’école oscille entre le formatif, les enseignants souhaitant former les élèves et le bilan. (Antibi 2003, Baillat, Ketele, Paquay, Thelot 2008).

cet « examen blanc » que se basent les orientations et passages en seconde.

En fait, comme le rappelle A. Antibi, **la note est sélective** dans les matières considérées comme importantes pour l'orientation. Elle n'est pas sélective lorsqu'il n'y a pas d'enjeux sociaux, par exemple dans les disciplines considérées comme mineures (dessin, musique, sport[5]), dans les grandes écoles (la sélection est faite à l'entrée) ou dans les lycées professionnels (il n'y a plus de sélection).

2.1. La constante macabre

Les enseignants sont donc conscients de faire de la sélection même lorsqu'ils affirment que leur rôle est de former. **Les très nombreuses expériences en docimologie montrent qu'une même copie peut obtenir entre 0 et 20**, même lorsque la compétence testée est précisée et même en mathématiques[6].

Les enseignants apprennent généralement à concevoir des contrôles avec 5 points de questions de cours (la « question cadeau »), 5 points d'applications directes du cours, 5 points d'applications plus complexes et enfin, un problème plus difficile « pour occuper les meilleurs ». L'objectif est généralement qu'aucun élève ne s'ennuie, et qu'un élève qui ait « fait des efforts » obtienne « la moyenne ». De cela découle, ce que A. Antibi a appelé « la constante macabre » à savoir que, dans chaque classe, le niveau des contrôles est adapté afin que la moyenne soit à peu près constante. De sorte que, des élèves n'ayant pas la moyenne en classe de première S au lycée Louis-Le-Grand et non autorisés à passer en classe supérieure obtiennent le baccalauréat S mention très bien après une terminale dans un établissement moins coté. Ce qui laisse penser que leurs notes à Louis-le-Grand n'étaient pas révélatrices de leur niveau mais bien du fait qu'ils étaient moins performants que leurs camarades de classe. Les enseignants ont souvent un rapport mitigé à la note. D'une part, c'est un outil pour « mettre les élèves au travail », mais d'autre part ils craignent de donner une impression de moindre exigence si la note est « trop bonne » ; c'est-à-dire que le devoir aurait été « trop facile ».

Ainsi, la note reste stable dans un groupe, du moins tant que les notes ont vocation à sélectionner (Antibi 2003).

Les notes sont souvent modifiées par les évaluations ; quand l'évaluation devient une façon de prouver l'efficacité d'une politique, elle est transformée (Duterq 2001).

Enfin, l'évaluation est un des éléments de la motivation des élèves. A. Antibi relate notamment comment, lorsque des enseignants changent leurs modalités d'évaluation, cela améliore les relations en classe ; impossible en revanche d'en savoir plus sur la hausse réelle du niveau des élèves.

« **L'existence de « la constante macabre » permet bien sûr de comprendre l'invariance de l'échec scolaire malgré les réformes successives** »[7]. De fait, puisque la note moyenne est constante dans une classe, et même si elle peut fluctuer individuellement légèrement dans l'année, le taux d'échec reste, également constant.

A. Antibi va jusqu'à provoquer : « La note augmente lorsque les profs n'essaient plus de pratiquer la constante macabre : que cela dit-il du niveau ? »

2.2. Les notations externes (PISA, le bac…)

Pour combattre « la constante macabre », **l'évaluation standardisée semble une solution**. La note étant surtout une manière de classer et d'orienter et finalement rarement une façon de définir un niveau, il y a un intérêt réel aux évaluations externes, (Mons Crachay 2011, p84) : « il s'agit d'administrer, à des cohortes d'élèves fréquentant les classes d'un certain niveau du cursus une (ou des) épreuves(s) portant sur les connaissances et compétences scolaires, épreuves dont l'élaboration, l'administration et la correction sont homogénéisées ». Ces évaluations ont commencé dès la fin du XIXème siècle au Canada et depuis les années 1960 en Europe même si leur développement est particulièrement marqué depuis les années 2000.

cet « examen blanc » que se basent les orientations et passages en seconde.

En fait, comme le rappelle A. Antibi, la note est sélective dans les matières considérées comme importantes pour l’orientation. Elle n’est pas sélective lorsqu’il n’y a pas d’enjeux sociaux, par exemple dans les disciplines considérées comme mineures (dessin, musique, sport[5]), dans les grandes écoles (la sélection est faite à l’entrée) ou dans les lycées professionnels (il n’y a plus de sélection).

2.1. La constante macabre

Les enseignants sont donc conscients de faire de la sélection même lorsqu’ils affirment que leur rôle est de former. Les très nombreuses expériences en docimologie montrent qu’une même copie peut obtenir entre 0 et 20, même lorsque la compétence testée est précisée et même en mathématiques[6].

Les enseignants apprennent généralement à concevoir des contrôles avec 5 points de questions de cours (la « question cadeau »), 5 points d’applications directes du cours, 5 points d’applications plus complexes et enfin, un problème plus difficile « pour occuper les meilleurs ». L’objectif est généralement qu’aucun élève ne s’ennuie, et qu’un élève qui ait « fait des efforts » obtienne « la moyenne ». De cela découle, ce que A. Antibi a appelé « la constante macabre » à savoir que, dans chaque classe, le niveau des contrôles est adapté afin que la moyenne soit à peu près constante. De sorte que, des élèves n’ayant pas la moyenne en classe de première S au lycée Louis-Le-Grand et non autorisés à passer en classe supérieure obtiennent le baccalauréat S mention très bien après une terminale dans un établissement moins coté. Ce qui laisse penser que leurs notes à Louis-le-Grand n’étaient pas révélatrices de leur niveau mais bien du fait qu’ils étaient moins performants que leurs camarades de classe. Les enseignants ont souvent un rapport mitigé à la note. D’une part, c’est un outil pour « mettre les élèves au travail », mais d’autre part ils craignent de donner une impression de moindre exigence si la note est « trop bonne » ; c'est-à-dire que le devoir aurait été « trop facile ».

Ainsi, la note reste stable dans un groupe, du moins tant que les notes ont vocation à sélectionner (Antibi 2003).

Les notes sont souvent modifiées par les évaluations ; quand l’évaluation devient une façon de prouver l’efficacité d’une politique, elle est transformée (Duterq 2001).

Enfin, l’évaluation est un des éléments de la motivation des élèves. A. Antibi relate notamment comment, lorsque des enseignants changent leurs modalités d’évaluation, cela améliore les relations en classe ; impossible en revanche d’en savoir plus sur la hausse réelle du niveau des élèves.

« L’existence de « la constante macabre » permet bien sûr de comprendre l’invariance de l’échec scolaire malgré les réformes successives »[7]. De fait, puisque la note moyenne est constante dans une classe, et même si elle peut fluctuer individuellement légèrement dans l’année, le taux d’échec reste, également constant.

A. Antibi va jusqu’à provoquer : « La note augmente lorsque les profs n’essaient plus de pratiquer la constante macabre : que cela dit-il du niveau ? »

2.2. Les notations externes (PISA, le bac...)

Pour combattre « la constante macabre », l'évaluation standardisée semble une solution. La note étant surtout une manière de classer et d'orienter et finalement rarement une façon de définir un niveau, il y a un intérêt réel aux évaluations externes, (Mons Crachay 2011, p84) : « il s'agit d'administrer, à des cohortes d'élèves fréquentant les classes d'un certain niveau du cursus une (ou des) épreuves(s) portant sur les connaissances et compétences scolaires, épreuves dont l'élaboration, l'administration et la correction sont homogénéisées ». Ces évaluations ont commencé dès la fin du XIXème siècle au Canada et depuis les années 1960 en Europe même si leur développement est particulièrement marqué depuis les années 2000.

2.2.1. Pour ce qui est des examens, le scepticisme règne

Pour le baccalauréat par exemple, l'année 2013 fut selon des médias « surnotés » (d'après l'Express[8] et le Figaro[9] notamment) tandis que les commentaires en ligne exprimaient largement la défiance généralisée face à « un examen que tout le monde a réussi ». Avec 90% d'obtention, nombre de personnes s'interrogent : « faut-il supprimer le bac ?», « est-ce encore un examen ? » « C'est comme le tour de France, on le dope ». Impossible, évidemment, de dire si le niveau a réellement augmenté ou non. Un lycéen[10] m'avait dit après avoir eu son bac en 2006 :

« J'ai eu mon bac grâce au CPE : avec un mois de cours en moins pendant les émeutes [de novembre 2005], deux mois de grève avec le CPE, sur quoi pouvaient-ils nous interroger ? » Lycéen, 19 ans, 2006.

La question se pose depuis plus de trente ans, laissant planer le **doute sur la progression réelle des performances scolaires** des élèves tandis qu'en sport, personne n'imagine que la vitesse des champions ne s'est pas améliorée. De fait, en course à pied, on mesure toujours un temps pour une distance parcourue tandis qu'à l'école**, les savoirs transmis ont évolué**, rendant impossible la comparaison entre le grand-père qui connaissait le nom des plantes et des départements de France et le petit-fils qui sait surfer sur Internet depuis ses deux ans.

2.2.2. PISA, Programme International de Suivi des Acquis

PISA est un ensemble d'études internationales menées par l'OCDE, sur la lecture, les mathématiques et les sciences. Sous forme de QCM, son évaluation est « automatique ». Il intéresse les chercheurs mais **la mesure de compétence ne correspond pas nécessairement à ce sur quoi est mis l'accent dans le système français**.

En effet, PISA évalue l'aptitude à réaliser des tâches qui s'inscrivent dans des situations de la vie réelle et qui dépendent d'une compréhension approfondie de concepts fondamentaux, et non l'acquisition de connaissances spécifiques[11]. Or, les élèves français ont une tendance à être plus perturbés par les épreuves « concrètes » que par les exercices abstraits (Antibi 2003) et donc par les présupposés de PISA.

« Par exemple, s'agissant de la compréhension de l'écrit, ces points forts et ces points faibles sont très liés à la conception de l'enseignement de la lecture qui est la nôtre. En effet, la France se positionne plutôt au-dessus de la moyenne internationale dans les compétences « s'informer » (…) et « interpréter » (…) les élèves français obtiennent un score inférieur à la moyenne internationale dans la compétence « réagir » » [12] Cette dernière compétence n'est traditionnellement pas ou peu enseignée, explique l'auteur de l'article. Le même auteur explique que le protocole PISA ne peut pas rendre compte des programmes français. Point de vue partagé par la plupart des associations professionnelles d'enseignants.

Plusieurs études visant spécifiquement à défendre les programmes français se sont penchées sur le rapport des enfants français aux QCM, examen peu pratiqué dans l'hexagone, pour montrer en quoi cela était un désavantage pour PISA par rapport aux autres pays.

De l'importance donc, d'aller mesurer autre chose, et notamment des données plus qualitatives.

PISA a cependant orienté le socle commun des compétences jugées indispensables pour vivre au XXIème siècle.

2.2.3. Les évaluations externes permettent de s'interroger

Les évaluations externes ne sont donc pas la panacée.

Il y a souvent une confusion dans les types de notes externalisées qui servent à la fois à l'évaluation d'un système et, selon les cas : à accéder à la suite du parcours d'études (baccalauréat, examens), à obtenir un

2.2.1. Pour ce qui est des examens, le scepticisme règne

Pour le baccalauréat par exemple, l'année 2013 fut selon des médias « surnotés » (d'après l'Express[8] et le Figaro[9] notamment) tandis que les commentaires en ligne exprimaient largement la défiance généralisée face à « un examen que tout le monde a réussi ». Avec 90% d'obtention, nombre de personnes s'interrogent : « faut-il supprimer le bac ?», « est-ce encore un examen ? » « C'est comme le tour de France, on le dope ». Impossible, évidemment, de dire si le niveau a réellement augmenté ou non. Un lycéen[10] m'avait dit après avoir eu son bac en 2006 :

« J'ai eu mon bac grâce au CPE : avec un mois de cours en moins pendant les émeutes [de novembre 2005], deux mois de grève avec le CPE, sur quoi pouvaient-ils nous interroger ? »
Lycéen, 19 ans, 2006.

La question se pose depuis plus de trente ans, laissant planer le doute sur la progression réelle des performances scolaires des élèves tandis qu'en sport, personne n'imagine que la vitesse des champions ne s'est pas améliorée. De fait, en course à pied, on mesure toujours un temps pour une distance parcourue tandis qu'à l'école, les savoirs transmis ont évolué, rendant impossible la comparaison entre le grand-père qui connaissait le nom des plantes et des départements de France et le petit-fils qui sait surfer sur Internet depuis ses deux ans.

2.2.2. PISA, Programme International de Suivi des Acquis

PISA est un ensemble d'études internationales menées par l'OCDE, sur la lecture, les mathématiques et les sciences. Sous forme de QCM, son évaluation est « automatique ». Il intéresse les chercheurs mais la mesure de compétence ne correspond pas nécessairement à ce sur quoi est mis l'accent dans le système français.

En effet, PISA évalue l'aptitude à réaliser des tâches qui s'inscrivent dans des situations de la vie réelle et qui dépendent d'une compréhension approfondie de concepts fondamentaux, et non l'acquisition de connaissances spécifiques[11]. Or, les élèves français ont une tendance à être plus perturbés par les épreuves « concrètes » que par les exercices abstraits (Antibi 2003) et donc par les présupposés de PISA.

« Par exemple, s'agissant de la compréhension de l'écrit, ces points forts et ces points faibles sont très liés à la conception de l'enseignement de la lecture qui est la nôtre. En effet, la France se positionne plutôt au-dessus de la moyenne internationale dans les compétences « s'informer » (...) et « interpréter » (...) les élèves français obtiennent un score inférieur à la moyenne internationale dans la compétence « réagir » » [12] Cette dernière compétence n'est traditionnellement pas ou peu enseignée, explique l'auteur de l'article. Le même auteur explique que le protocole PISA ne peut pas rendre compte des programmes français. Point de vue partagé par la plupart des associations professionnelles d'enseignants.

Plusieurs études visant spécifiquement à défendre les programmes français se sont penchées sur le rapport des enfants français aux QCM, examen peu pratiqué dans l'hexagone, pour montrer en quoi cela était un désavantage pour PISA par rapport aux autres pays.

De l'importance donc, d'aller mesurer autre chose, et notamment des données plus qualitatives.

PISA a cependant orienté le socle commun des compétences jugées indispensables pour vivre au XXIème siècle.

2.2.3. Les évaluations externes permettent de s'interroger

Les évaluations externes ne sont donc pas la panacée.

Il y a souvent une confusion dans les types de notes externalisées qui servent à la fois à l'évaluation d'un système et, selon les cas : à accéder à la suite du parcours d'études (baccalauréat, examens), à obtenir un

PISA a notamment permis à la France de s'interroger sur le redoublement, et d'identifier les établissements qui allaient être en réseau ambition réussite.

« Au total, plus qu'un outil de publicité de l'action publique, d'aide à la décision ou d'évaluation des réformes, **l'évaluation standardisée s'impose de fait comme un instrument de reconfiguration**, de transformation de l'action publique en éducation, visant à faire changer les pratiques pédagogiques et évoluer les rapports de pouvoir entre les acteurs traditionnels dans ce secteur, l'Etat, les professionnels de l'éducation que sont les enseignantes et les personnels d'encadrement, les collectivités locales et les parents » (Mons Crahay 2011).

A ce titre, l'évaluation externe est un outil précieux, à conserver. Il faut cependant rester conscient qu'elle est loin de permettre une évaluation fiable et exhaustive des dispositifs visant à améliorer les performances scolaires des élèves, comme nous allons le voir.

3. La note est-elle toujours le résultat tangible ?

3.1 Le facteur temps

La présente section vise à montrer que l'amélioration des résultats scolaires est un des derniers résultats par ordre de survenance dans un certain nombre de configurations, bien après l'amélioration de l'estime de soi.

Les données présentées proviennent des travaux de recherche d'Hélène Vandelle lors de sa thèse.

H Vandelle a fait passer aux mêmes élèves de lycée professionnel, en octobre 2007 et en mai 2008 le test ETES (échelle toulousaine de l'estime de soi) de N Oubrayrie. Ce test est composé de 60 questions et donne une évaluation sur 5 dimensions

- Le soi émotionnel
- Le soi social
- Le soi scolaire, mesure des représentations que les élèves ont de leur performance scolaire
- Le soi physique
- Le soi futur

L'estime de soi est mesurée à partir de l'ensemble des dimensions.

Les élèves, ayant une image d'eux très dégradée au départ, ont également répondu à une vingtaine de questions permettant de construire un indicateur du « sentiment d'efficacité personnel ». Cet indicateur présente le sentiment de compétences, le rapport à l'effort et la persistance dans l'effort. Il s'agit d'une mesure de perception de contrôle. Il pourrait se traduire par « je vais y arriver ».

Les questionnaires exploitables (correctement remplis aux deux dates) sont au nombre de 135 élèves de 11 classes dans 7 lycée professionnels de l'académie de Versailles. Avec 98 filles et 37 garçons.

PISA a notamment permis à la France de s'interroger sur le redoublement, et d'identifier les établissements qui allaient être en réseau ambition réussite.

« Au total, plus qu'un outil de publicité de l'action publique, d'aide à la décision ou d'évaluation des réformes, l'évaluation standardisée s'impose de fait comme un instrument de reconfiguration, de transformation de l'action publique en éducation, visant à faire changer les pratiques pédagogiques et évoluer les rapports de pouvoir entre les acteurs traditionnels dans ce secteur, l'Etat, les professionnels de l'éducation que sont les enseignantes et les personnels d'encadrement, les collectivités locales et les parents » (Mons Crahay 2011).

A ce titre, l'évaluation externe est un outil précieux, à conserver. Il faut cependant rester conscient qu'elle est loin de permettre une évaluation fiable et exhaustive des dispositifs visant à améliorer les performances scolaires des élèves, comme nous allons le voir.

3. La note est-elle toujours le résultat tangible ?

3.1 Le facteur temps

La présente section vise à montrer que l'amélioration des résultats scolaires est un des derniers résultats par ordre de survenance dans un certain nombre de configurations, bien après l'amélioration de l'estime de soi.

Les données présentées proviennent des travaux de recherche d'Hélène Vandelle lors de sa thèse.

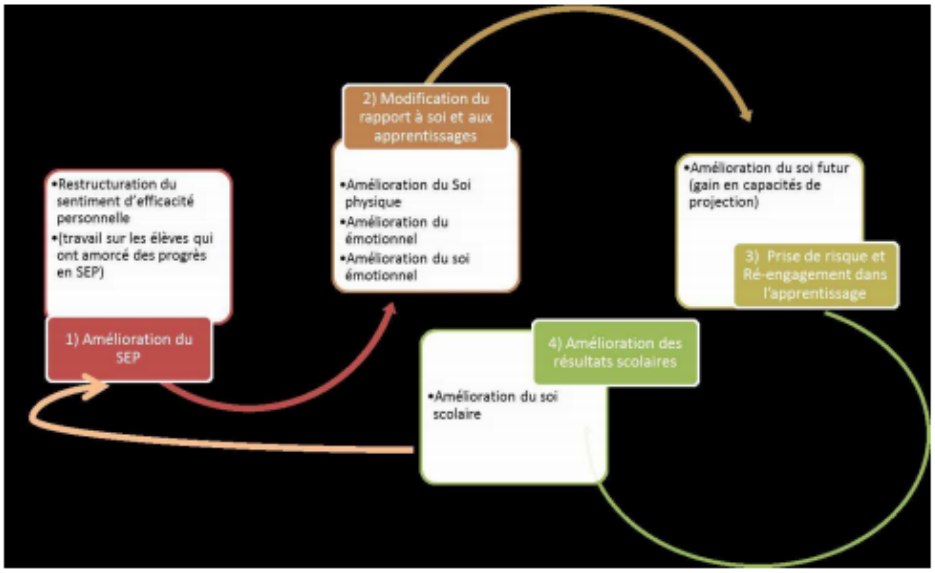
H Vandelle a fait passer aux mêmes élèves de lycée professionnel, en octobre 2007 et en mai 2008 le test ETES (échelle toulousaine de l'estime de soi) de N Oubrayrie. Ce test est composé de 60 questions et donne une évaluation sur 5 dimensions

Le soi émotionnel Le soi social Le soi scolaire, mesure des représentations que les élèves ont de leur performance scolaire Le soi physique Le soi futur

L'estime de soi est mesurée à partir de l'ensemble des dimensions.

Les élèves, ayant une image d'eux très dégradée au départ, ont également répondu à une vingtaine de questions permettant de construire un indicateur du « sentiment d'efficacité personnel ». Cet indicateur présente le sentiment de compétences, le rapport à l'effort et la persistance dans l'effort. Il s'agit d'une mesure de perception de contrôle. Il pourrait se traduire par « je vais y arriver ».

Les questionnaires exploitables (correctement remplis aux deux dates) sont au nombre de 135 élèves de 11 classes dans 7 lycée professionnels de l'académie de Versailles. Avec 98 filles et 37 garçons.



L'amélioration des résultats scolaires n'arrive qu'en dernière étape. En effet, le chemin de progression (qui dépend du genre de l'adolescent) passe par le renforcement du sentiment de sécurité (être protégé, faire confiance) et d'identité (savoir ce dont on est capable ou pas) puis le sentiment d'appartenance au groupe, puis le sentiment de détermination (se fixer des buts) et enfin le sentiment de compétence (atteindre ses buts). Ce n'est que lorsque le sentiment de compétence est restauré que les progrès en matière de résultats scolaires deviennent visibles, au minimum après un travail de plusieurs mois voire souvent de plusieurs années. Si l'on considère que, pour les élèves étudiés, entre le début et la fin de l'année, une majorité de garçons notamment n'avaient progressé que sur la dimension « soi physique », l'amélioration du « soi scolaire » qui correspond à une auto-évaluation de la progression en termes de résultats restait très en retard. Peut-on dire que les élèves n'ont pas progressé ?

Inversement, parmi les élèves en progrès sur la dimension soi-futur, le soi scolaire a plus progressé[13] que pour les autres élèves. Cette étape est préalable aux résultats.

En conséquence, il est important de relativiser « l'inefficacité d'un dispositif » lorsque les notes ne s'améliorent pas. Le temps est souvent long et l'aspect scolaire est, parfois, le dernier progrès, suivant une série de progrès qu'il convient de ne pas négliger.

3.2 Le risque d'une évaluation fausse

Dans un excellent article, Pascal Bressoux et Laurent Lima[14] montrent comment une évaluation de l'influence de la taille des classes sur les performances des élèves, pourtant réalisée sérieusement avec des évaluations externes standardisées, aurait pu se révéler lourdement contre-productive.

Le ministère de l'éducation nationale envisageait d'augmenter la taille des classes en s'appuyant sur une étude qui montrait que « la diminution des effectifs dans les classes n'a pas d'effets avérés sur les résultats des élèves et que les très petites écoles ne s'avèrent pas toujours plus performantes[15] »

L’amélioration des résultats scolaires n’arrive qu’en dernière étape. En effet, le chemin de progression (qui dépend du genre de l’adolescent) passe par le renforcement du sentiment de sécurité (être protégé, faire confiance) et d’identité (savoir ce dont on est capable ou pas) puis le sentiment d’appartenance au groupe, puis le sentiment de détermination (se fixer des buts) et enfin le sentiment de compétence (atteindre ses buts). Ce n’est que lorsque le sentiment de compétence est restauré que les progrès en matière de résultats scolaires deviennent visibles, au minimum après un travail de plusieurs mois voire souvent de plusieurs années. Si l’on considère que, pour les élèves étudiés, entre le début et la fin de l’année, une majorité de garçons notamment n’avaient progressé que sur la dimension « soi physique », l’amélioration du « soi scolaire » qui correspond à une auto-évaluation de la progression en termes de résultats restait très en retard. Peut-on dire que les élèves n’ont pas progressé ?

Inversement, parmi les élèves en progrès sur la dimension soi-futur, le soi scolaire a plus progressé[13] que pour les autres élèves. Cette étape est préalable aux résultats.

En conséquence, il est important de relativiser « l’inefficacité d’un dispositif » lorsque les notes ne s’améliorent pas. Le temps est souvent long et l’aspect scolaire est, parfois, le dernier progrès, suivant une série de progrès qu’il convient de ne pas négliger.

3.2 Le risque d’une évaluation fausse

Dans un excellent article, Pascal Bressoux et Laurent Lima[14] montrent comment une évaluation de l’influence de la taille des classes sur les performances des élèves, pourtant réalisée sérieusement avec des évaluations externes standardisées, aurait pu se révéler lourdement contre-productive.

Le ministère de l’éducation nationale envisageait d’augmenter la taille des classes en s’appuyant sur une étude qui montrait que « la diminution des effectifs dans les classes n’a pas d’effets avérés sur les résultats des élèves et que les très petites écoles ne s’avèrent pas toujours plus performantes[15] »

Elle s'était déroulée de la manière suivante : au cours de l'année 2002-2003, en ZEP, une centaine de classes de CP ont vu leurs effectifs réduits à 10 élèves par classe avec autant de classes témoins qui sont restées autour de 20 élèves.

Les élèves des classes à effectif réduit et des classes témoins ont passé des épreuves standardisées en octobre 2002, en mars 2003 et en juin 2003. Le niveau initial des deux groupes était comparable en octobre 2002. **Mais, entre octobre 2002 et juin 2003, les élèves ont obtenu de meilleurs résultats dans les classes à effectifs normaux que dans celles à effectifs réduits**. D'où la proposition du ministère d'augmenter la taille des classes.

Bressoux et Lima ont étudié, pour les classes expérimentales et les classes témoins, l'ancienneté des enseignants en CP et l'ancienneté des enseignants dans la profession. Le processus d'affectation des enseignants est resté inconnu des chercheurs mais il semble qu'il ait été décidé d'aider les jeunes enseignants en leur confiant ces classes « allégées » : **les enseignants étaient significativement plus jeunes et moins expérimentés dans les classes à effectifs réduits que dans les classes témoins**.

Un modèle statistique intégrant ces variables montre que la variable qui prédomine dans l'explication de la note est, de très loin, l'ancienneté des enseignants en CP, concordant, en cela, avec toutes les études sur le sujet. Une fois neutralisée cette variable, on retrouve probablement que, à expérience égale des enseignants, les classes de plus petits effectifs présentent bien de meilleurs résultats (les mesures restent incertaines car les échantillons deviennent trop faibles).

En tout état de cause, les résultats de cette étude basée sur les notes sont faux et dangereux : un paramètre important ayant été occulté. La réduction à l'analyse des notes peut entrainer des biais importants.

3.3. Les aspects qualitatifs

Le dispositif « Cours le matin, sport l'après-midi » vise à améliorer les performances scolaires des élèves en profitant des apports des études sur les rythmes scolaires. Le principe qui a conduit à expérimenter cette organisation de l'emploi du temps est que les journées sont trop longues pour un apprentissage en continu, que les pics de concentration se situent le matin et qu'au contraire vers 15h, il est complexe de mobiliser l'attention des élèves.

Les établissements qui ont participé à l'expérience étaient volontaires et recevaient 5000€ par an pour financer les surcoûts dans la mise en place de partenariats sportifs ou culturels.

Une évaluation qualitative de ce dispositif a été commanditée par le Fond d'Expérimentation pour la Jeunesse a été publiée en 2012.

Le rapport propose notamment, en annexe[16], le compte-rendu d'entretien suivant, que j'ai réalisé avec Laure, 12 ans, élève en classe de 4ème à propos de l'expérience vécue en 5ème.

Laure vit avec sa mère, femme au foyer, sa grande sœur de 14 ans, qui est au collège en 3ème et 5 demi-frères et sœurs plus jeunes. Son géniteur qui a reconnu sa sœur aînée ne l'a pas reconnue elle.

Elle est inscrite à l'UNSS pour la gymnastique depuis la 6ème et aime dans l'ensemble le sport.

Elle n'aime pas trop l'école car « il faut travailler » mais elle s'entend bien avec ses pairs et ses professeurs. C'est elle, cependant, qui a craché sur d'autres personnes lors d'une sortie.

Elle était considérée par ses enseignants comme « difficile » même si l'enseignante de français pense que c'est « du gâchis » lié à des difficultés familiales.

Elle s'était déroulée de la manière suivante : au cours de l'année 2002-2003, en ZEP, une centaine de classes de CP ont vu leurs effectifs réduits à 10 élèves par classe avec autant de classes témoins qui sont restées autour de 20 élèves.

Les élèves des classes à effectif réduit et des classes témoins ont passé des épreuves standardisées en octobre 2002, en mars 2003 et en juin 2003. Le niveau initial des deux groupes était comparable en octobre 2002. Mais, entre octobre 2002 et juin 2003, les élèves ont obtenu de meilleurs résultats dans les classes à effectifs normaux que dans celles à effectifs réduits. D'où la proposition du ministère d'augmenter la taille des classes.

Bressoux et Lima ont étudié, pour les classes expérimentales et les classes témoins, l'ancienneté des enseignants en CP et l'ancienneté des enseignants dans la profession. Le processus d'affectation des enseignants est resté inconnu des chercheurs mais il semble qu'il ait été décidé d'aider les jeunes enseignants en leur confiant ces classes « allégées » : les enseignants étaient significativement plus jeunes et moins expérimentés dans les classes à effectifs réduits que dans les classes témoins.

Un modèle statistique intégrant ces variables montre que la variable qui prédomine dans l'explication de la note est, de très loin, l'ancienneté des enseignants en CP, concordant, en cela, avec toutes les études sur le sujet. Une fois neutralisée cette variable, on retrouve probablement que, à expérience égale des enseignants, les classes de plus petits effectifs présentent bien de meilleurs résultats (les mesures restent incertaines car les échantillons deviennent trop faibles).

En tout état de cause, les résultats de cette étude basée sur les notes sont faux et dangereux : un paramètre important ayant été occulté. La réduction à l'analyse des notes peut entrainer des biais importants.

3.3. Les aspects qualitatifs

Le dispositif « Cours le matin, sport l'après-midi » vise à améliorer les performances scolaires des élèves en profitant des apports des études sur les rythmes scolaires. Le principe qui a conduit à expérimenter cette organisation de l'emploi du temps est que les journées sont trop longues pour un apprentissage en continu, que les pics de concentration se situent le matin et qu'au contraire vers 15h, il est complexe de mobiliser l'attention des élèves.

Les établissements qui ont participé à l'expérience étaient volontaires et recevaient 5000€ par an pour financer les surcoûts dans la mise en place de partenariats sportifs ou culturels.

Une évaluation qualitative de ce dispositif a été commanditée par le Fond d'Expérimentation pour la Jeunesse a été publiée en 2012.

Le rapport propose notamment, en annexe[16], le compte-rendu d'entretien suivant, que j'ai réalisé avec Laure, 12 ans, élève en classe de 4

ème

à propos de l'expérience vécue en 5

ème

.

Laure vit avec sa mère, femme au foyer, sa grande sœur de 14 ans, qui est au collège en 3ème et 5 demi-frères et sœurs plus jeunes. Son géniteur qui a reconnu sa sœur aînée ne l'a pas reconnue elle.

Elle est inscrite à l'UNSS pour la gymnastique depuis la 6ème et aime dans l'ensemble le sport.

Elle n'aime pas trop l'école car « il faut travailler » mais elle s'entend bien avec ses pairs et ses professeurs. C'est elle, cependant, qui a craché sur d'autres personnes lors d'une sortie.

Elle était considérée par ses enseignants comme « difficile » même si l'enseignante de français pense que c'est « du gâchis » lié à des difficultés familiales.

Elle s'est sentie plus fatiguée, n'arrivait pas à se lever et avait envie de rester au lit (ce qu'elle faisait d'ailleurs régulièrement). Ses résultats scolaires se sont dégradés et elle a moins travaillé l'an dernier. ***Elle pense que cette année elle a plus d'énergie et qu'elle arrive mieux à travailler.*** *Il semble que l'année passée, difficile sur un plan physique et moral, l'ait amené à améliorer son hygiène de vie et qu'elle puisse en tirer quelques bénéfices avec un décalage d'un an.*

Laure n'a certes pas progressé en classe de 5ème, loin de là. Le dispositif était encore très loin de réussir à produire les effets escomptés. Néanmoins, partant d'une jeune fille qui se couchait à 3h du matin et réussir à faire en sorte qu'elle prenne l'habitude sur un an de dormir à 21h ou 22h est une réelle amélioration de son sommeil. Si Laure avait bénéficié de cet aménagement sur plusieurs années, on peut imaginer que, dans la durée, les performances scolaires auraient fini par s'améliorer. En tout état de cause, il serait dommage de conclure de l'absence de résultats scolaires que le dispositif ne fonctionne pas.

En fait d'autres chercheurs (Brito 2012 ; Cousin 1998) montrent que les résultats d'une expérimentation sont rarement réductibles à la note et qu'il faut intégrer les données qualitatives.

O. Brito (2012) montre notamment dans une analyse des écoles alternatives par rapport à des écoles « traditionnelles » qu'aucun résultat n'est visible sur les notes. En revanche, on peut observer des conséquences du passage par des pédagogies différentes sur le bien-être des élèves.

A. Grisay expliquait : « Il est quelque peu désolant de constater que les indicateurs relatifs à la concertation, à la cohérence des pratiques et des objectifs, au style de direction, à la pédagogie innovante, aux ressources matérielles, à la formation continue ne paraissent pas avoir le moindre lien avec la performance des élèves. »[17]

4. Conclusion : Comment évaluer ?

4.0. L'intérêt de l'évaluation est réel

Le besoin d'évaluation des dispositifs expérimentaux est réel. Le besoin d'objectivation aussi[18]. Les chiffres semblent souvent, aux commanditaires plus « parlant » et plus « objectifs » que des commentaires ou des analyses qualitatives. Ils offrent un accès rapide et simplifié, voire simpliste, à la réalité. Comme nous l'avons vu, les notes sont un aspect de la mesure des performances des élèves souffrant de nombreux biais. Ces évaluations peuvent être utilisées, du moment que l'on est conscient de leurs limites et qu'elles n'exonèrent pas d'autres modes d'évaluation des dispositifs d'amélioration des performances scolaires des élèves.

Les normes d'évaluation diffèrent d'un système d'acteurs à un autre : on repère les normes de la société française d'évaluation, les normes administratives, les normes des chercheurs en science de l'éducation, etc. Cela correspondant à des paradigmes différents.

Quelles que soient les normes, on peut néanmoins admettre le besoin d'évaluation.

L'évaluation, quand elle est comporte plusieurs points de vue, permet d'une part d'observer les éventuels effets d'un dispositif sur les performances scolaires, mais aussi de garder une capacité de découverte d'effets secondaires inattendus. De la même manière qu'un médicament peut être plus intéressant pour ses effets secondaires que pour ceux prévus, c'est en cherchant des informations qualitatives que nous nous sommes rendu-compte que le premier effet de « cours le matin, sport l'après-midi » était l'amélioration du bien-être des élèves : l'évaluation doit permettre de mieux comprendre en quoi le dispositif agit.

Enfin, comme l'explique Y. Dutercq (2001) l'évaluation est un mode de mobilisation pour l'innovation : « Les évaluations permettent aux innovateurs de se remobiliser régulièrement et de rompre avec la lassitude qui survient dans tout exercice de longue durée ; elles interpellent ceux qui n'y prennent pas part et les font mettre

Elle s’est sentie plus fatiguée, n’arrivait pas à se lever et avait envie de rester au lit (ce qu’elle faisait d’ailleurs régulièrement). Ses résultats scolaires se sont dégradés et elle a moins travaillé l’an dernier. Elle pense que cette année elle a plus d’énergie et qu’elle arrive mieux à travailler. Il semble que l’année passée, difficile sur un plan physique et moral, l’ait amené à améliorer son hygiène de vie et qu’elle puisse en tirer quelques bénéfices avec un décalage d’un an.

Laure n’a certes pas progressé en classe de 5ème, loin de là. Le dispositif était encore très loin de réussir à produire les effets escomptés. Néanmoins, partant d’une jeune fille qui se couchait à 3h du matin et réussir à faire en sorte qu’elle prenne l’habitude sur un an de dormir à 21h ou 22h est une réelle amélioration de son sommeil. Si Laure avait bénéficié de cet aménagement sur plusieurs années, on peut imaginer que, dans la durée, les performances scolaires auraient fini par s’améliorer. En tout état de cause, il serait dommage de conclure de l’absence de résultats scolaires que le dispositif ne fonctionne pas.

En fait d’autres chercheurs (Brito 2012 ; Cousin 1998) montrent que les résultats d’une expérimentation sont rarement réductibles à la note et qu’il faut intégrer les données qualitatives.

O. Brito (2012) montre notamment dans une analyse des écoles alternatives par rapport à des écoles « traditionnelles » qu’aucun résultat n’est visible sur les notes. En revanche, on peut observer des conséquences du passage par des pédagogies différentes sur le bien-être des élèves.

A. Grisay expliquait : « Il est quelque peu désolant de constater que les indicateurs relatifs à la concertation, à la cohérence des pratiques et des objectifs, au style de direction, à la pédagogie innovante, aux ressources matérielles, à la formation continue ne paraissent pas avoir le moindre lien avec la performance des élèves. »[17]

4. Conclusion : Comment évaluer ?

4.0. L’intérêt de l’évaluation est réel

Le besoin d’évaluation des dispositifs expérimentaux est réel. Le besoin d’objectivation aussi[18]. Les chiffres semblent souvent, aux commanditaires plus « parlant » et plus « objectifs » que des commentaires ou des analyses qualitatives. Ils offrent un accès rapide et simplifié, voire simpliste, à la réalité. Comme nous l’avons vu, les notes sont un aspect de la mesure des performances des élèves souffrant de nombreux biais. Ces évaluations peuvent être utilisées, du moment que l’on est conscient de leurs limites et qu’elles n’exonèrent pas d’autres modes d’évaluation des dispositifs d’amélioration des performances scolaires des élèves.

Les normes d’évaluation diffèrent d’un système d’acteurs à un autre : on repère les normes de la société française d’évaluation, les normes administratives, les normes des chercheurs en science de l’éducation, etc. Cela correspondant à des paradigmes différents.

Quelles que soient les normes, on peut néanmoins admettre le besoin d'évaluation.

L'évaluation, quand elle est comporte plusieurs points de vue, permet d'une part d'observer les éventuels effets d'un dispositif sur les performances scolaires, mais aussi de garder une capacité de découverte d'effets secondaires inattendus. De la même manière qu'un médicament peut être plus intéressant pour ses effets secondaires que pour ceux prévus, c'est en cherchant des informations qualitatives que nous nous sommes rendu-compte que le premier effet de « cours le matin, sport l'après-midi » était l'amélioration du bien-être des élèves : l'évaluation doit permettre de mieux comprendre en quoi le dispositif agit.

Enfin, comme l'explique Y. Dutercq (2001) l'évaluation est un mode de mobilisation pour l'innovation : « Les évaluations permettent aux innovateurs de se remobiliser régulièrement et de rompre avec la lassitude qui survient dans tout exercice de longue durée ; elles interpellent ceux qui n'y prennent pas part et les font mettre en question leur pratique »[19].

mesures influenceront l'organisation de l'établissement et le processus d'apprentissage. Néanmoins, diverses études montrent l'intérêt de l'auto-évaluation des dispositifs innovants par les établissements eux-mêmes, en évaluation « embarquée » c'est-à-dire avec un accompagnement des ajustements au fur et à mesure que se déroule l'expérience. L'évaluation assume alors sa fonction « d'outil de reconfiguration d'un dispositif » et perd celle de « rendre compte à un public ».

Dans un article de 2001[20], Anton Strittmatter résume le résultat d'une expérience de quatre ans menée dans une trentaine d'établissements Suisse. Le « système d'évaluation formative de la qualité » ou SEFOQ a été testé dans 25 écoles-pilotes allant de la maternelle au collège puis a été étendu dans trois autres établissements dont un lycée. Le principe est qu' « il appartient à chaque établissement qui intègre le système d'autoévaluation d'assumer et de vérifier les exigences de qualités qui lui sont propres ».

Le SEFOQ vise à impliquer l'ensemble des acteurs (élèves, parents, enseignants, gestionnaires, politiques…). Pour y parvenir, les évaluations par des instances ayant un pouvoir hiérarchique sur l'école ou les enseignants sont proscrites, de même que tout usage « publicitaire » ou instrumentalisation de la démarche. Ainsi l'évaluation interdit ou limite la publication des classements pour limiter les effets de compétition.

Deux actions majeures sont mises en œuvre dans le cadre des activités d'évaluation : une partie recherche-développement pour améliorer la régulation du système, intervenir en cas de crise, etc. et une partie évaluation à proprement parler incluant des données statistiques, éventuellement des enquêtes, mais pas de notes d'élèves. Les sources de ces évaluations sont : les (auto)observations, l'évaluation par les pairs, l'expertise, les retours des parents, de la direction, et de tous les acteurs.

L'autoévaluation conduit à la mise en œuvre du projet d'école, des outils, des analyses de pratiques, de la formation continue pour les enseignants, des décisions en terme d'équipement, de modes de travail, d'organisation mais aussi à résoudre des difficultés spécifiques identifiées. Des sous-groupes « qualité » sont mandatés pour réfléchir sur des problèmes spécifiques.

Le système recommande une évaluation externe et une régulation à intervalle régulier (tous les cinq ans).

Les expériences permettent du soutien entre les membres (en particulier dans les sous-groupes), une économie en termes d'efforts, partage des divers points de vue (contrôle et intégration), et une légitimité des indicateurs choisis. Les écoles qui ont le mieux réussi le passage à l'autoévaluation sont celles qui ont été accompagnées, au moins par un comité de pilotage externe pour limiter l'instrumentalisation et la méfiance à l'égard de la direction.

4.2. Les diagnostics ou l'auto-évaluation des jeunes plutôt que les évaluations sommatives

Des écoles innovantes (notamment le Collège Lycée Elitaire Pour Tous de Grenoble ou le Lycée Autogéré de Paris) pratiquent l'auto-évaluation des élèves par eux-mêmes : les lycéens déterminent à intervalle régulier des critères pour estimer s'ils ont progressé et se notent sur ces critères. Il est cependant nécessaire que ces évaluations ne donnent pas lieu à publicité pour rester honnête.

La première piste pour évaluer un dispositif est d'utiliser d'autres indicateurs que les performances scolaires et en particulier les tests d'estime de soi ou les questionnaires psycho-sociaux. En effet, comme l'ont montré les travaux d'H. Vandelle, l'amélioration de l'estime de soi précède les progrès scolaires. Il est donc légitime de les étudier en priorité.

Les diagnostics sont également une piste intéressante. D'une part les jeunes ne se mobilisent pas de la même façon pour une évaluation diagnostique et pour une évaluation sommative. D'autre part, en tenant compte des

mesures influenceront l'organisation de l'établissement et le processus d'apprentissage. Néanmoins, diverses études montrent l'intérêt de l'auto-évaluation des dispositifs innovants par les établissements eux-mêmes, en évaluation « embarquée » c'est-à-dire avec un accompagnement des ajustements au fur et à mesure que se déroule l'expérience. L'évaluation assume alors sa fonction « d'outil de reconfiguration d'un dispositif » et perd celle de « rendre compte à un public ».

Dans un article de 2001[20], Anton Strittmatter résume le résultat d'une expérience de quatre ans menée dans une trentaine d'établissements Suisse. Le « système d'évaluation formative de la qualité » ou SEFOQ a été testé dans 25 écoles-pilotes allant de la maternelle au collège puis a été étendu dans trois autres établissements dont un lycée. Le principe est qu' « il appartient à chaque établissement qui intègre le système d'autoévaluation d'assumer et de vérifier les exigences de qualités qui lui sont propres ».

Le SEFOQ vise à impliquer l'ensemble des acteurs (élèves, parents, enseignants, gestionnaires, politiques...). Pour y parvenir, les évaluations par des instances ayant un pouvoir hiérarchique sur l'école ou les enseignants sont proscrites, de même que tout usage « publicitaire » ou instrumentalisation de la démarche. Ainsi l'évaluation interdit ou limite la publication des classements pour limiter les effets de compétition.

Deux actions majeures sont mises en œuvre dans le cadre des activités d'évaluation : une partie recherche- développement pour améliorer la régulation du système, intervenir en cas de crise, etc. et une partie évaluation à proprement parler incluant des données statistiques, éventuellement des enquêtes, mais pas de notes d'élèves. Les sources de ces évaluations sont : les (auto)observations, l'évaluation par les pairs, l'expertise, les retours des parents, de la direction, et de tous les acteurs.

L'autoévaluation conduit à la mise en œuvre du projet d'école, des outils, des analyses de pratiques, de la formation continue pour les enseignants, des décisions en terme d'équipement, de modes de travail, d'organisation mais aussi à résoudre des difficultés spécifiques identifiées. Des sous-groupes « qualité » sont mandatés pour réfléchir sur des problèmes spécifiques.

Le système recommande une évaluation externe et une régulation à intervalle régulier (tous les cinq ans).

Les expériences permettent du soutien entre les membres (en particulier dans les sous-groupes), une économie en termes d'efforts, partage des divers points de vue (contrôle et intégration), et une légitimité des indicateurs choisis. Les écoles qui ont le mieux réussi le passage à l'autoévaluation sont celles qui ont été accompagnées, au moins par un comité de pilotage externe pour limiter l'instrumentalisation et la méfiance à l'égard de la direction.

4.2. Les diagnostics ou l'auto-évaluation des jeunes plutôt que les évaluations sommatives

Des écoles innovantes (notamment le Collège Lycée Elitaire Pour Tous de Grenoble ou le Lycée Autogéré de Paris) pratiquent l'auto-évaluation des élèves par eux-mêmes : les lycéens déterminent à intervalle régulier des critères pour estimer s'ils ont progressé et se notent sur ces critères. Il est

cependant nécessaire que ces évaluations ne donnent pas lieu à publicité pour rester honnête.

La première piste pour évaluer un dispositif est d'utiliser d'autres indicateurs que les performances scolaires et en particulier les tests d'estime de soi ou les questionnaires psycho-sociaux. En effet, comme l'ont montré les travaux d'H. Vandelle, l'amélioration de l'estime de soi précède les progrès scolaires. Il est donc légitime de les étudier en priorité.

Les diagnostics sont également une piste intéressante. D'une part les jeunes ne se mobilisent pas de la même façon pour une évaluation diagnostique et pour une évaluation sommative. D'autre part, en tenant compte des résultats mis en évidence par d'autres chercheurs (notamment Gérard 2008, et Brézillon et Champault 2008),